

Prosodie de langues à tons (naxi et vietnamien), prosodie de l'anglais : éclairages croisés

Alexis Michaud

▶ **To cite this version:**

Alexis Michaud. Prosodie de langues à tons (naxi et vietnamien), prosodie de l'anglais : éclairages croisés. Linguistique. Université de la Sorbonne nouvelle - Paris III, 2005. Français. ⟨NNT : ⟩. ⟨tel-00130149⟩

HAL Id: tel-00130149

https://theses.hal.science/tel-00130149v1

Submitted on 9 Feb 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

UNIVERSITE PARIS 3 – SORBONNE NOUVELLE

ECOLE DOCTORALE 268 « LANGAGE ET LANGUES »

Prosodie de langues à tons (naxi et vietnamien), prosodie de l'anglais : éclairages croisés

thèse de Doctorat nouveau régime

présentée par Pierre-Alexis MICHAUD

discipline : Phonétique

Directrice de thèse : Mme Jacqueline VAISSIÈRE

soutenue le 12 décembre 2005 devant le jury composé de :

M. Claude DELMAS, Examinateur

M. Klaus KOHLER, Prérapporteur

Mme Martine MAZAUDON, Prérapporteur

Mme Aliyah MORGENSTERN, Examinatrice

M. Mario ROSSI, Prérapporteur

Mme Jacqueline VAISSIÈRE, Directrice

Volume 1

Remerciements

Parmi les nombreuses personnes dont l'aide a permis la réalisation de ce travail, je souhaite remercier tout particulièrement : David Bradley pour ses indications sur les langues yi, et pour des commentaires sur un projet de publication ; mes professeurs de linguistique Pierre Cotte, Laurent Danon-Boileau et Claude Delmas ; Denis Creissels pour des commentaires sur la perspective de typologie prosodique proposée dans la thèse ; Jean-Yves Dommergues pour ses conseils dans le domaine statistique ; Michel Ferlus pour ses explications en phonétique historique ; Nathalie Henrich pour son aide en acoustique ; Daniel Hirst pour la formulation du projet ; Monique Hoa pour des conseils bibliographiques ; Larry Hyman pour ses commentaires sur deux projets de publication ; Klaus Kohler pour ses conseils, à Paris, Kiel et Nara ; Shinji Maeda pour des scripts MatLab ; Martine Mazaudon et Boyd Michailovsky pour leurs conseils dans l'analyse phonologique du naxi et leur relecture de travaux, la communication de données de langue naxi, et la formation à l'étude des langues tibéto-birmanes ; Aliyah Morgenstern pour ses conseils et ses encouragements ; Christine Mooshammer pour la relecture de projets de publications ; Francis Nolan pour son accueil au laboratoire de phonétique de Cambridge, et Geoffrey Potter pour son aide dans la réalisation des enregistrements dans ce même laboratoire ; Bernard Roubeau pour les mesures de débit d'air ; Laurent Sagart pour d'utiles commentaires en phonétique historique ; Michel Scheffers pour ses indications au sujet des mesures de formants ; Jackson Sun pour d'utiles échanges sur la prosodie de langues tibéto-birmanes ; Michel Viel pour son soutien ; Vũ-Ngọc Tuấn pour la réalisation de programmes, pour son aide en acoustique et en programmation, et pour le travail en commun sur la langue vietnamienne.

Mes vifs remerciements à tous les membres du Laboratoire Phonétique et Phonologie (LPP), en particulier Nick Clements, Lise Crevier-Buchman, François Dell, Bernard Gautheron, Barbara Kühnert, Jean-Léo Léonard et Annie Rialland. Mes remerciements tout particuliers à Jacqueline Vaissière, pour ses critiques aussi bienveillantes qu'énergiques, et plus généralement son aide de tous les instants. Parmi les doctorants du LPP et d'ailleurs, merci à Angélique Amelot, Laura Bennetts, Marc Brunelle, Olivier Corbin, Cédric Gendrot, Guillaume Jacques, Takeki Kamiyama, Aimée Lahaussois, Liberty Lidz, Julie Montagu, David Mortensen, Oliver Niebuhr, Frédérique Passot, Cédric Patin, Patricia Pieri, Olivier Piot, Martine Toda, Alice Vittrant.

Ma gratitude va également à Huang Xing (黄行教授) pour l'invitation à mener des recherches en Chine ; à Kong Jiangping (孔江平教授) pour l'autorisation d'utiliser le matériel de son laboratoire ; à Xu Lin (徐琳女士) pour les données de langue bai qu'elle a aimablement communiquées ; aux consultants et amis naxi : Guo Dalie (郭大烈教授), He Jiren (和即仁先生), He Jiezhen (和洁珍女士), He Xueguang (和学光教授), He Xuan (和漩), He Qin (和沁), He Yong (和涌), He Xixian (和希贤), Mu Ruhua (木汝华), et leurs familles ; et aux consultants, collègues et amis vietnamiens, en particulier Trần Trí Dõi (Université Nationale de Hanoi) et Vũ Xuân Xuyến (Université Paris 7).

Il va de soi que les erreurs d'interprétation des informations qui m'ont été communiquées sont de mon fait.

Du matériel d'enregistrement a été gracieusement prêté par le laboratoire Langues et Civilisations à Tradition Orale (LACITO, CNRS-Universités Paris 3 et Paris 4). Une partie des enregistrements électroglottographiques a été réalisée au Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI, CNRS-Universités Paris-Sud et Pierre et Marie Curie) et au Laboratoire d'Acoustique Musicale (LAM, CNRS-Université Pierre et Marie Curie-Ministère de la Culture et de la Communication) ; mes vifs remerciements aux Directeurs de ces laboratoires.

Les missions en Chine ont été financées grâce au concours de l'Université Paris 3 (Projet innovant 2002-2003), du Ministère de l'Education Nationale et de la Recherche (programme Aires Culturelles) et de la Fondation de France.

Conventions concernant les citations en langues étrangères

La traduction en français des citations en anglais n'est pas apparue nécessaire. Les citations en chinois ont été traduites par nous[1].

Mise en page et conventions typographiques[2]

Dans l'idée de faciliter la lecture, trois formes de mise en valeur ont été employées : **les caractères gras mettent en relief les idées principales**,

> les paragraphes à double bordure gauche signalent les articulations de l'exposé,

des paragraphes encadrés contiennent des résumés en tête de chapitre et des bilans d'étape dans le cours du texte.

Par commodité technique, les figures sont disposées sur des planches séparées.

Les références sont appelées sous le format ＜auteur＞ ＜année de publication＞ ; par souci de clarté, le titre de la publication est parfois ajouté dans le corps du texte : par exemple (Rialland 1998, « Systèmes prosodiques africains : une source d'inspiration majeure pour les théories phonologiques multilinéaires ») au lieu de la simple référence (Rialland 1998). Une annexe à la bibliographie présente les ouvrages en chinois, la version actuelle du logiciel EndNote ne permettant pas leur insertion dans la bibliographie principale.

---

[1] Les noms chinois ont été transcrits en suivant la notation *pinyin* pour les noms de personnes de Chine continentale. Certains auteurs originaires de Taïwan ou Hong Kong signent d'un nom occidentalisé (Jackson Sun pour 孫天心), ou romanisé en suivant le système anglais Wade-Giles (Shih Chilin pour 石基琳), avec parfois des variantes (Chao Yuen-Ren pour 趙元任 ; la transcription de la seconde syllabe, *yuen*, est une variante courante de la transcription standard *yuan*) ; le choix de l'auteur est alors suivi dans les références bibliographiques. Les noms japonais ont été présentés dans l'ordre prénom-nom : par exemple Haruo Kubozono (窪薗晴夫), Shinji Maeda (前田眞治).

[2] Les chiffres à décimales sont notés selon la convention anglo-saxonne uniformément employée par les langages informatiques : par exemple, pour ½, 0.5 et non 0,5. En conséquence, les nombres à plus de trois chiffres n'ont pas été divisés en milliers, millions etc. par des points : pour éviter les confusions, 44100 (quarante-quatre mille cent) n'est pas noté 44.100.

Table des matières

Une table des matières détaillée figure en tête de chaque chapitre. Une table des matières complète (en 11 pages) figure tout à la fin du second volume.

Introduction ____ ***1***

1. Problématique de recherche : l'hypothèse d'un *partage des ressources* entre l'accentuation et l'intonation ____ 1

2. Enjeu au plan de la modélisation prosodique et de la typologie prosodique ____ 2

3. Présentation du cadre global de recherche ____ 4

4. Notions centrales de la modélisation prosodique ____ 10

Chapitre I. Méthode expérimentale : enjeux et outils ____ ***17***

1. Intérêt d'une mesure de la qualité de voix (et de la fréquence fondamentale) par électroglottographie ____ 17

2. Remarques sur les fréquences formantiques et l'intensité ____ 29

Chapitre II. Aperçu d'ensemble de l'intonation du naxi ____ ***41***

Premier volet : Recherches sur l'*intonation syntaxique* de la langue naxi ____ **41**

1. Déclinaison et frontières : réflexion sur les notions, et observations préliminaires ____ 42

2. Réduplication et prosodie : données synchroniques et hypothèse diachronique concernant les formes réduplicatives en naxi ____ 52

3. Approche de la ligne de déclinaison et des frontières par l'emploi d'énoncés dont toutes les syllabes portent le même ton ____ 74

Tableau de l'intonation du naxi, deuxième volet : Recherches sur l'intonation pragmatique de la langue naxi ____ **85**

1. Moyens grammaticaux de structuration de l'information ____ 86

2. Réalisation d'une intensification par des modifications intonatives locales ____ 89

Chapitre III. Etude comparée des modifications locales induites par l'intonation pragmatique en naxi, vietnamien et anglais ____ ***107***

1. Données anglaises ____ 111

2. Données naxi ____ 128

3. Données vietnamiennes ____ 143

4. Synthèse des résultats : comparaison entre les trois langues ____ 161

Chapitre IV. Discussion générale ____ ***167***

1. L'hypothèse d'un partage des ressources entre accentuation et intonation ____ 167

2. Implications théoriques des observations : choix pour une modélisation prosodique ____ 178

3. Réflexion typologique ____ 191

Conclusion ____ ***225***

1. Conclusions au plan de la technique expérimentale ____ 225

2. Conclusions concernant la prosodie des trois langues étudiées ____ 226

3. Conclusions au plan de la modélisation prosodique et de la typologie ____ 229

Glossaire et explication des abréviations 233
Références 249
Annexe 1. Observations sur les phonèmes du naxi 297
1. Les initiales 305
2. Observations synchroniques sur les rimes 329
3. Prononciation des phonèmes de la variété locale de chinois 345
4. Phénomènes à l'échelle du mot : fusion syllabique et intégration nominale 357
5. Variations intonatives dans la réalisation des phonèmes 365
6. Aperçu de la variation dialectale : FK, NL 367
Eléments de conclusion 378
Annexe 2. Résultats expérimentaux complémentaires, et outils informatiques 381
1. Données anglaises complémentaires 381
2. Résultats de mesures de débit d'air des tons vietnamiens B2, D1 et D2 391
3. Etude de DECPA, amplitude des pics de fermeture sur le signal électroglottographique 399
4. Scripts développés dans l'environnement MATLAB 404
Annexe 3. Perspectives théoriques complémentaires 423
1. L'alignement des points-clefs de la courbe de fréquence fondamentale avec la « ligne des segments » : la problématique et sa remise en cause 423
2. Tons et structure métrique en chinois 430
Figures du volume I 435
Graphiques 491
Figures des Annexes 494
Cartes 514

Index 516
Table des figures du volume I 518
Table des photographies et des cartes 520
Table des graphiques 520
Table des figures des Annexes 521
Table des matières détaillée *fin du volume*

Table des matières de l'Introduction

1. Problématique de recherche : l'hypothèse d'un *partage des ressources* entre l'accentuation et l'intonation 1

2. Enjeu au plan de la modélisation prosodique et de la typologie prosodique 2

3. Présentation du cadre global de recherche 4

3.1. Complémentarité entre phonétique et phonologie 5

3.2. Variété des langues et synthèse typologique 6

3.3. Multiplicité des paramètres prosodiques, et utilité des explorations physiologiques 6

3.3.1. Rôle synchronique et diachronique de la qualité de voix dans les systèmes tonals 8

3.3.2. Association fréquente d'une certaine qualité de voix à des phénomènes intonatifs 9

4. Notions centrales de la modélisation prosodique 10

4.1. Cadre britannique 10

4.2. Cadre de la présente étude 11

4.2.1. Intonation syntaxique (= composante syntaxique de l'intonation) 12

4.2.2. Intonation pragmatique (= composante pragmatique de l'intonation) 13

Bilan de l'Introduction : Questions centrales 14

Introduction

La présente recherche se veut un éclairage croisé entre prosodie de langues à tons et prosodie de l'anglais, et plus généralement une contribution à la typologie prosodique.

Les termes *prosodie* et *intonation* sont discutés plus avant dans l'Introduction, section 4 ; pour anticiper, la prosodie inclut l'*accentuation*, l'*intonation* (qui elle-même comporte une composante syntaxique et une composante pragmatique, et inclut également l'expression des *émotions et attitudes*), et les *facteurs de performance* dont le rythme et le débit de parole. L'accentuation telle qu'elle est entendue ici inclut seulement les phénomènes distinctifs lexicaux, à l'exclusion de l'accent démarcatif (« accent fixe »), lequel, dans ce jeu de définitions, appartient à la composante syntaxique de l'intonation, et de l'accent d'insistance, qui appartient à la composante pragmatique de l'intonation. Les phénomènes distinctifs au plan lexical qui constituent l'accentuation sont (selon les langues) l'accent lexical, les accents mélodiques, les tons lexicaux, ou les registres de qualité de voix[1].

1. Problématique de recherche : l'hypothèse d'un *partage des ressources* entre l'accentuation et l'intonation

La comparaison entre langues permet de soumettre à vérification l'hypothèse d'un *partage des ressources* entre phénomènes lexicaux et intonation. Cette hypothèse formulée par divers auteurs prédit par exemple qu'une langue où l'emploi de la longueur est contraint au niveau lexical par la présence d'oppositions de longueur vocalique et de consonnes géminées (comme le finnois) emploiera moins la longueur au plan intonatif, tandis qu'une langue qui ne connaît pas de telles oppositions aura toutes chances d'employer l'allongement dans le marquage des frontières intonatives (au sujet de cette notion, dont les équivalents anglais sont *intonational boundaries* et *intonational junctures*, voir ci-dessous, section 4.2.2), et de façon expressive.

Dans le cas des langues considérées, ce raisonnement prédit un emploi intonatif moindre de la fréquence fondamentale en naxi et vietnamien, langues qui possèdent des tons, qu'en

[1] Inclure dans l'*accentuation* les tons lexicaux (qui existent dans des langues comme le naxi et le vietnamien) et les registres de qualité de voix (qui existent dans des langues comme le môn) a quelque chose de paradoxal, « accentuation » évoquant immédiatement *accent*, et non *ton*, encore moins le phénomène peu courant des *registres de qualité de voix*. Ce choix résulte de l'extension par analogie de la catégorie « accentuation » du cadre proposé par M. Rossi et J. Vaissière ; il fait ressortir une caractéristique commune à l'accent lexical (accent libre), aux tons lexicaux et aux registres de qualité de voix, qui est de posséder un statut lexical distinctif. Il pourrait être indiqué de choisir un autre nom pour la catégorie en question ; cela n'a toutefois pas paru avisé au stade actuel du travail, car le choix d'une nouvelle dénomination aurait pu brouiller les correspondances de vocabulaire avec les cadres théoriques choisis pour référence.

anglais, où la présence d'un accent lexical ne préjuge pas du détail de la courbe de fréquence fondamentale de la syllabe qui porte cet accent. De même pour la qualité de voix (définie ici comme le mode de vibration des plis vocaux : la manière dont est généré le bourdonnement qui constitue le voisement) : en anglais et en naxi, langues dans lesquelles la qualité de voix n'est pas spécifiée lexicalement, il paraît intéressant de vérifier dans quelle mesure la qualité de voix est employée au plan intonatif, et de comparer ces résultats avec ceux obtenus pour le vietnamien, langue dans laquelle la qualité de voix entre dans la définition de certains tons[1].

2. Enjeu au plan de la modélisation prosodique et de la typologie prosodique

La comparaison des trois langues étudiées doit permettre de contribuer aux débats actuels dans le domaine des études intonatives, dans lesquels la notion de ton occupe une place centrale. Aux Etats-Unis, en particulier depuis les travaux de Kenneth Pike (Pike 1945), Zellig Harris (Harris 1951), George Trager et Henry Smith (Trager et Smith 1951), se développe **un courant qui défend le brillant paradoxe selon lequel l'intonation d'un énoncé anglais pourrait se décomposer en séquences de tons** (quatre niveaux tonals chez William Leben [Leben 1976, « The tones in English intonation »] ; trois chez Mark Liberman [Liberman 1975] ; deux chez Janet Pierrehumbert [Pierrehumbert 1980]). L'approche de Janet Pierrehumbert fait ressortir de façon originale la richesse des phénomènes intonatifs de l'anglais en s'imposant un inventaire restreint de catégories descriptives : essentiellement les tons binaires H (haut) et L (bas), des *frontières*, et une hiérarchie de *constituants prosodiques*. Cette approche a par la suite été étendue à un grand nombre de langues : un ouvrage collectif tout récent (*Prosodic typology: the phonology of intonation and phrasing*, dirigé par Sun-Ah Jun, Oxford University Press, 2005) présente des systèmes de notations de l'intonation de onze langues typologiquement variées (dont le chinois mandarin et le chinois cantonais), qui adoptent les principes de la notation ToBI *(Tones and Break Indices)* proposée pour l'anglais (« ToBI: A Standard for Labeling English Prosody », Silverman, Beckman *et al.* 1992). **Cette extension spectaculaire, jusqu'à la description de l'intonation de langues à tons lexicaux, stimule la réflexion typologique : dans quelle mesure les « tons intonatifs » sont-ils semblables à des tons lexicaux ?** Robert Ladd évoque la thèse extrême selon laquelle ils seraient identiques, dans la mesure où la prosodie de toutes les langues pourrait se décrire en termes de *séquences tonales* (Ladd 1992 ; une longue citation est reprise au chapitre IV, section 2.1.5).

[1] Pour plus de détails, voir le chapitre IV, en particulier sections 3.1.4 (au sujet des langues à registres de qualité de voix) et 3.1.5.3 (au sujet des tons qui combinent hauteur et qualité de voix).

Le présent travail offre l'occasion de contribuer à cette réflexion, partant de l'idée selon laquelle la comparaison des tons lexicaux présents dans certaines langues avec les « tons intonatifs » postulés par ce courant de recherche peut contribuer à éclaircir les perspectives typologiques. En effet, **les modèles évoqués ci-dessus, dits *autosegmentaux-métriques*** (Ladd 1996)**, empruntent au modèle *autosegmental*, élaboré pour rendre compte de faits observés dans des langues à tons lexicaux d'Afrique**[1]. Le choix de langues d'Asie paraît prometteur pour le renouvellement des problématiques de recherche en prosodie : en effet, dans certaines langues d'Afrique, la marge de jeu intonatif local laissée à l'énonciateur est réduite, comme en témoignent notamment les travaux de Yetunde Laniran (Laniran 1992, *Intonation in Tone Languages: the Phonetic Implementation of Tones in Yoruba*), Annie Rialland et Stéphane Robert (Rialland et Robert 2001, 2003, Rialland 2004). **A l'opposé, dans beaucoup de langues à tons d'Asie, des phénomènes intonatifs se superposent aux tons lexicaux d'une façon qui paraît particulièrement saillante** (comme le soulignaient déjà Chao Yuen-ren et Kenneth Pike : Chao Yuen-ren 1933, 1935, Pike 1948:29)[2]. L'étude de langues à tons d'Asie offrirait donc l'occasion d'observer en détail l'interaction des tons lexicaux avec l'intonation, et ainsi d'asseoir la modélisation prosodique sur des fondements typologiques consolidés. (Pour anticiper sur la Discussion, indiquons d'emblée que la version extrême des théories autosegmentales-métriques qu'expose Robert Ladd soulève des difficultés au plan typologique ; les observations réalisées sur les trois langues étudiées paraissent en revanche éminemment compatibles avec les modèles *superpositionnels* de la prosodie.)

Le présent travail vise également à contribuer à une meilleure connaissance de la prosodie des langues à tons d'Asie par la description du système tonal de trois dialectes du **naxi** (désignation chinoise : na^{4} xi^{1} yu^{3} 纳西语, autonyme : /nɑ̀ hi ¯/), **langue minoritaire de Chine qui présente trois tons ponctuels (haut H, moyen M, et bas L)**[3], et sa comparaison avec le

---

[1] Au sujet de l'histoire de l'influence que l'étude des systèmes prosodiques africains a exercée sur le développement des théories phonologiques *multilinéaires*, voir Rialland 1998.

[2] Ces travaux pionniers ont depuis été complétés par des études expérimentales, qui seront évoquées au fil de la discussion.

[3] Le naxi, génétiquement proche du sous-groupe birman-yi de la famille tibéto-birmane (Bradley 1975), est parlé par environ 300000 locuteurs principalement concentrés dans la Région Autonome Naxi de Ljiang (yun^{2}nan^{2} sheng^{3}, li^{4}jiang^{1} na^{4}xi^{1}zu^{2} zi^{4}zhi^{4}xian^{4} 云南省丽江纳西族自治县) de la province chinoise du Yunnan. Il est notamment décrit par Li Lincan, Zhang Kun *et al.* 1953 ; Mazaudon et Michailovsky 1979 ; He Jiren et Jiang Zhuyi 1985 ; Fang Guoyu et He Zhiwu 1995. Le naxi est un cas d'école de langue à tons ponctuels : dans cette langue, chaque syllabe porte un ton lexical haut (ˊ), moyen (¯), ou bas (ˋ) (ci-après notés H, M, L) ; il existe également un ton montant bas-haut (ˋˊ), plus rare (décrit en détail dans un article soumis pour publication). Le principal dialecte naxi étudié est le naxi *occidental* tel qu'il est parlé dans la ville de Lijiang (丽江) et aux alentours. Le principal point d'enquête est le village de /ɑ̄ ʂə̀/ (ci-après abrégé en AS), à environ 20 km au nord-est de la ville de Lijiang ; en chinois : jin^{1}shan^{1} xiang^{1}, wen^{2}hua^{4} xing^{2}zheng^{4}cun^{1}, leng^{3}bu^{4}luo^{2} zi^{4}ran^{2}cun^{1} 丽江，金山乡，文化行政村，冷不罗自然村. Sauf mention contraire, les données

vietnamien (parler de Hanoi), dont les tons se définissent par leur courbe de fréquence fondamentale mais aussi par des caractéristiques de qualité de voix (mode de vibration des plis vocaux). Le travail sur la langue naxi a occupé une place importante dans le calendrier de la thèse (deux séjours de trois mois sur le terrain ont été effectués, en 2002 et 2004) ; l'étude de l'intonation d'une langue nécessite, nous semble-t-il, une compréhension de l'intérieur. L'intonation entretient des relations complexes avec d'autres domaines de la langue, en particulier la morphosyntaxe (relations qui peuvent être de renforcement mutuel, d'affinité, ou d'équivalence, certains moyens morphosyntaxiques pouvant par exemple être remplacés par des procédés intonatifs dans un registre de langue familier). Au plan de la comparaison entre langues, il convient pareillement de prêter attention aux correspondances qui peuvent exister entre des processus intonatifs dans une langue et des processus morphosyntaxiques dans une autre. (Au chapitre III est ainsi fourni un exemple de phénomène déclaré « paralinguistique » en anglais, qui est morphosyntaxique en vietnamien et en naxi.) Le naxi est une langue relativement peu documentée ; les exigences du travail de terrain, qui confronte le chercheur aux questions les plus variées (phonologie, morphosyntaxe et sémantique, mais aussi dialectologie et sociolinguistique), coïncident avec notre souci d'une compréhension de l'intonation dans le cadre de la langue dans son entier. Notre contribution à la description du système phonémique du naxi est présentée dans l'Annexe 1 ; lexique, morphosyntaxe et textes suivis feront l'objet de publications ultérieures.

La section 3 expose le cadre général des recherches, la section 4, de façon plus ciblée, les notions centrales du cadre d'analyse prosodique choisi.

3. Présentation du cadre global de recherche[1]

Le cadre général de recherche est une conception de la phonétique-phonologie qui met l'accent sur la complémentarité entre diverses approches des faits de langues : phonétique et phonologie ; typologie et études de cas ; travail sur corpus construits et étude d'une parole plus naturelle ; étude de la production, de l'acoustique et de la perception ; étude qualitative et traitement automatique.

---

proviennent de ce dialecte. Le deuxième dialecte étudié est celui du village de /ndɑ̄ lè/ (abrégé ci-après en NL) : en chinois, jin[1]shan[1] xiang[1], gui[4]feng[1] da[4]lai[2] xing[2]zheng[4] xia[4] cun[1] 丽江, 金山乡, 贵峰大来行政下村/大来二村), à une douzaine de kilomètres au sud de la ville de Lijiang, dialecte qui présente certains schèmes tonals particuliers. Enfin, certaines données du parler du village de /fɣ̀ kʰō/ (abrégé en FK), feng[4]ke[1] xiang[1], shan[4]mei[3] xing[2]zheng[4]cun[1], heng[2]ke[3] er[4] zu[3] 丽江, 奉科乡, 善美行政村, 恒可二组, à la frontière entre le dialecte occidental et le dialecte oriental, sont également utilisées dans la discussion. Les cartes géographiques proposées dans l'Annexe 1 indiquent la localisation des points d'enquête.

[1] Le présent travail nous semble participer de l'esprit d'une équipe, le Laboratoire de Phonétique et Phonologie, et plus généralement de directions actuelles de recherche en phonétique-phonologie ; il va néanmoins de soi que l'exposé proposé dans cette section n'engage que son auteur.

3.1. Complémentarité entre phonétique et phonologie

La phonologie s'est dans une certaine mesure fondée en se dissociant de la phonétique ; dans les termes souvent cités de Trubetzkoy, cette dernière se trouve redéfinie comme « la science de la face matérielle des sons du langage humain » :

> La seule tâche de la phonétique est de répondre à la question : « comment prononce-t-on ceci et cela ? » (...) Les sons du langage que la phonétique doit étudier possèdent un grand nombre de particularités acoustiques et articulatoires qui pour le phonéticien sont toutes importantes, car c'est seulement en considérant toutes ces particularités qu'il peut répondre d'une manière précise à la question que pose la prononciation du son dont il s'agit. Mais pour le phonologue la plupart de ces particularités sont tout à fait accessoires, car elles ne fonctionnent pas comme marques distinctives des mots. Aussi les sons du phonéticien et les unités du phonologue ne se recouvrent pas. (Trubetzkoy 1939 [1976], *Grundzüge der Phonologie* ; traduit par Jean Cantineau : *Principes de phonologie*, éd. Klincksieck, Paris, 1949, pp. 10-12)

Cette rupture a été favorable au développement parallèle des deux disciplines, et à une spécialisation accrue des chercheurs. **Phonétique et phonologie n'en demeurent pas moins deux faces d'une même discipline.** Comme le souligne Mario Rossi (Rossi 1995), dès l'après-guerre, l'ouvrage de Roman Jakobson, Gunnar Fant et Morris Halle *Preliminaries to Speech Analysis* (Jakobson, Fant *et al.* 1952) inaugure une collaboration renouvelée entre phonétique expérimentale et phonologie. Le titre d'un ouvrage de Ohala et Jaeger, *Experimental Phonology* (Ohala et Jaeger 1986), résume le souhait de rapprocher les deux disciplines. Un même programme est porté par l'expression de « laboratory phonology », popularisée par la conférence du même nom[1].

La convergence entre phonétique et phonologie n'est certes pas complète[2]. Certains chercheurs sont plutôt attirés par la théorie même s'ils sont soucieux de la soumettre à l'épreuve des faits ; d'autres sont plutôt attirés par l'expérimentation, et y consacrent l'essentiel de leur énergie ; d'autres enfin, amateurs de langues, sont avant tout soucieux de comprendre le fonctionnement synchronique et diachronique des faits qu'ils observent, sans nécessairement entrer dans les discussions théoriques ni dans le domaine expérimental. Cette différence tient en définitive au tempérament du chercheur (constat qui s'applique sans doute aux domaines scientifiques les plus variés) :

> Bien sûr, l'idéal affiché d'objectivité et de rigueur que s'imposent la linguistique et d'autres sciences humaines rend absolument taboue la question du tempérament, de

[1] Voir en particulier l'Introduction aux Actes de la première de cette série de conférences annuelles : Kingston et Beckman 1990.

[2] Voir les perspectives tracées par Stephen Anderson (Anderson 1985), et, dans le domaine français, la discussion des idées de Laks et Plénat 1993 par Di Cristo 2004:73-76, 84

> l'imaginaire, du désir du chercheur, qui seraient au mieux la part privée de la recherche, vouée à s'effacer devant les résultats, et personne n'a tenté encore une caractérologie des théories. Pourtant nul ne s'investit dans l'étude du langage sans y apporter un projet inconscient et nombre de querelles relèvent de la psychologie tout autant que de l'épistémologie (...). (Cotte 1993:9)

Il n'est guère concevable d'envisager la réunification en une discipline unique des sciences phonétiques, dont la pluralité est reconnue depuis 1932 dans le nom même du *Congrès des Sciences Phonétiques* qui rassemble périodiquement phonéticiens, phonologues, ingénieurs spécialistes de la communication parlée, psychologues, enseignants de langues et de chant, médecins, orthophonistes et d'autres encore. En revanche, **le dialogue continu entre chercheurs de sensibilités différentes (spécialistes de phonologie, de phonétique, de reconstruction des langues) et le travail en équipe permettent de pallier les limites de chacun**.

3.2. Variété des langues et synthèse typologique

La spécialité géographique de chacun, de par la structure des langues concernées, façonne dans une certaine mesure ses orientations théoriques : les spécialistes de langues africaines sont également spécialistes de la phonologie non linéaire et des tons ponctuels ; les spécialistes de langues asiatiques sont familiers de langues dans lesquelles les tons ont partie liée avec les segments, possèdent souvent des corrélats variés (hauteur, qualité de voix, articulation consonantique et vocalique), et s'associent aux syllabes selon des modalités différentes des tons subsahariens[1], de sorte que les problématiques de recherche s'orientent dans une direction différente. Les échanges avec des spécialistes de diverses langues paraissent donc prometteurs pour un progrès de la réflexion, au plan de la théorie générale aussi bien qu'au plan typologique.

3.3. Multiplicité des paramètres prosodiques, et utilité des explorations physiologiques

Ilse Lehiste a montré l'importance des trois paramètres prosodiques que sont la fréquence fondamentale, la durée et l'intensité (voir en particulier Lehiste 1970, *Suprasegmentals*). Dans les termes proposés par Mario Rossi :

> L'accentuation et l'intonation se manifestent dans un espace prosodique à trois dimensions. Ces dimensions objectives sont la fréquence fondamentale, l'intensité et le temps. Etant donné que l'important pour l'auditeur est ce qu'il entend et ce qu'il perçoit, les traits prosodiques doivent être fondés sur des dimensions définies à partir de la perception qu'on en a, ou qu'on est supposé en avoir ; ces dimensions

[1] Cette question est abordée en détail dans la Discussion (chapitre IV), section 3.1.

> perceptivement fondées sont respectivement la mélodie, la sonie, et pour le temps, l'allongement. (Rossi 1999:7 ; voir également Di Cristo 2004:82)

La facilité de la mesure de F_0 depuis l'avènement de l'informatique, jointe à l'importance évidente de ce paramètre au plan linguistique, a quelque peu occulté la variété des phénomènes intonatifs. L'intonation est fréquemment associée au seul paramètre de fréquence fondamentale[1]. Plusieurs systèmes de transcription font le choix de ne retenir que F_0 (notamment ToBI, et INTSINT, présenté dans l'Introduction de l'ouvrage collectif dirigé par D. Hirst et A. Di Cristo : Hirst et Di Cristo 1998). Ce choix présente un avantage de simplicité qui peut être décisif pour la didactique de l'intonation, ainsi que pour l'étiquetage de larges corpus en vue de leur traitement automatique. En revanche, il peut paraître réducteur dans une perspective de recherche fondamentale.

Dans le présent travail, les corrélats phonétiques de la prosodie sont *toutes les variations non prévisibles à partir de la connaissance des phonèmes et des faits d'accentuation* (définition commentée plus bas, section 1.2). Cet élargissement reflète le souci de n'exclure d'emblée aucune dimension des faits prosodiques, et d'attirer l'attention vers les mécanismes, objets d'analyse, plutôt que de la focaliser sur les paramètres, qui ne sont que des moyens d'aborder la réalité étudiée.

Ainsi, l'articulation des consonnes et des voyelles a part à la prosodie. L'évolution diachronique divergente qu'ont connue les syllabes sous accent et les syllabes inaccentuées dans de nombreuses langues témoigne de l'effet de la prosodie sur l'articulation des consonnes et les voyelles (voir par exemple Martinet 1955, et les hypothèses de Vaissière 2001, « Changements de sons et changements prosodiques : du latin au français »). Au plan synchronique, l'incidence de l'accent sur l'articulation est manifeste dans les langues comme l'anglais et le russe (« ... ce qui est phonologiquement pertinent pour l'accentuation russe du *mot* c'est la différence qualitative et quantitative entre les voyelles des syllabes accentuées et inaccentuées » : Trubetzkoy 1939 [1976]:243)[2] ; dans les langues où les variantes allophoniques des voyelles en syllabe accentuée et non accentuée n'appartiennent pas à des catégories phonémiques différentes, les différences n'en sont pas moins mesurables, comme en témoignent les travaux déjà relativement anciens cités par Ivan Fónagy (Fónagy 1983:110) ; des travaux expérimentaux récents sont consacrés aux relations entre l'articulation et des facteurs prosodiques tels que l'accent lexical, l'insistance intonative, le marquage des frontières (voir notamment Erickson, Fujimura *et al.* 2002, Mooshammer et Fuchs 2002, Cho 2005 et références citées, ainsi que Stone et Vatikiotis-Bateson 1995).

Dans le présent travail est plus particulièrement prise en compte la *qualité de voix*, définie ici comme le mode de vibration des plis vocaux : la manière dont est généré le

---

[1] Voir par exemple Fant 1974:5 : « The time variation of F_0 is the physical basis of intonation. »

[2] Signalons également l'étude expérimentale de Padgett et Tabain 2005, et références citées.

bourdonnement qui constitue le voisement. Dans la perspective qui vient d'être exposée, la qualité de voix trouve une place dans le champ des études intonatives. Elle correspond à la composante *glottique* de la production de la parole. (La figure 0.2, reprise de Catford 1977:18, permet de visualiser la décomposition du conduit vocal en trois composantes, sous-glottique, glottique et supraglottique ; cette question sera abordée à nouveau en section 3.) Une introduction à l'étude de la qualité de voix est proposée par John Catford, qui propose des exercices pour se familiariser avec divers mécanismes phonatoires. Un continuum relie les extrêmes de la qualité de voix (tels que voix tendue *[tense]*, soufflée *[breathy]*, craquée *[creaky]*) : ainsi, Catford nomme « voix chuchotée » *(whispery voice)* une voix légèrement soufflée, proposant l'exercice suivant pour la production de la voix murmurée : « ... starting with normal, relaxed, low frequency voice, and then, while it is going on, relaxing the vocal folds further till continuous whisper is heard mixed in with voice » (Catford 1977:99 ; un tableau des possibilités phonatoires du larynx humain est proposé par Laver 1980).

La qualité de voix joue des rôles variés d'une langue à l'autre (un panorama des emplois linguistiques de la qualité de voix est proposé par Gordon et Ladefoged 2001). Deux courants alimentent plus particulièrement la curiosité à l'endroit de la *qualité de voix*, et orientent la méthode d'analyse employée ici.

3.3.1. Rôle synchronique et diachronique de la qualité de voix dans les systèmes tonals

Dans l'apparition des tons du vietnamien, les registres de qualité de voix ont très certainement représenté une étape intermédiaire dans la transphonologisation d'oppositions consonantiques en oppositions tonales (processus bien connu, dans ses grandes lignes, depuis les travaux fondateurs de Maspero et Haudricourt [Maspero 1912, Haudricourt 1954, 1961, 1965] ; les principaux stades du développement des tons en vietnamien sont présentés au chapitre IV, section 3.1.5.3). **De façon très simplifiée, le schéma typique est le suivant : une opposition consonantique se transforme en une opposition de qualité de voix, puis en opposition tonale.** Les *langues à registres de qualité de voix* se situent à la seconde étape de ce processus : dans ces langues, le mode de vibration des plis vocaux est un paramètre distinctif au plan lexical. Ainsi, le môn (langue de la famille austroasiatique) possède deux registres, qui opposent des syllabes réalisées en phonation modale, aussi appelée phonation *claire* (non marquée ; ni particulièrement tendue ni particulièrement relâchée/soufflée) à des syllabes de même composition phonémique réalisées en phonation *soufflée (breathy voice)*[1]. Des systèmes de registre complexes sont décrits par Therapan Thongkum dans des parlers môn-khmers (Thongkum 1988) : une même syllabe peut porter deux qualités de voix, l'une en début de rime, l'autre en fin de rime (par exemple, voix soufflée suivie d'une glottalisation ; voir chapitre IV, section 3.1.4). Les oppositions de registres de voix peuvent

[1] John Catford *(op. cit.)* préfère parler de phonation *murmurée*, c'est-à-dire une phonation *légèrement soufflée*, réservant le terme « soufflé » pour une qualité de voix plus extrême, où le débit d'air est très élevé.

par la suite se transphonologiser en oppositions vocaliques, occasionnant une partition du système des voyelles (comme c'est le cas en cambodgien), ou une partition du système tonal : Abramson et ses collaborateurs rapportent qu'au seuil d'une enquête expérimentale sur les registres de la langue suai, ils ont constaté que l'opposition de qualité de voix était devenue une opposition tonale dans les dialectes étudiés (Abramson, Thongkum *et al.* 2004).

Sur cette base, **il demeure de nombreuses questions à traiter, pour parvenir à une compréhension affinée, en synchronie, des systèmes existants, et pour aborder la reconstruction d'états anciens des langues d'Asie à la lumière de ces phénomènes**[1].

Le choix d'étudier les variations intonatives dans une langue où la qualité de voix joue un rôle distinctif paraît offrir une perspective originale au sujet des relations entre hauteur et qualité de voix : **l'image du ton qui se dégage de sa variation allophonique suggère certaines possibilités d'évolution** ; une observation synchronique fine peut aider à l'intelligence de la diachronie.

3.3.2. Association fréquente d'une certaine qualité de voix à des phénomènes intonatifs

De même que la fréquence fondamentale, l'intensité, la durée, et d'autres paramètres, la qualité de voix tient une place dans le domaine intonatif, à divers niveaux (voir par exemple Beckman et Pierrehumbert 1992:397).

Qualité de voix et frontières : Des études expérimentales confirment la fréquence de l'association d'une certaine qualité de voix (typiquement une glottalisation, ou un passage en voix soufflée—voire un dévoisement complet) avec les frontières intonatives, dans des langues variées (voir par exemple, au sujet de l'anglais : Dilley et Shattuck-Hufnagel 1996 et Epstein 2002 ; au sujet du français, Fagyal 1995 ; au sujet du persan : Assadi 2003).

Qualité de voix et mise en valeur pragmatique : La qualité de voix a également part à la réflexion sur les corrélats physiologiques de l'*insistance* (mise en valeur pragmatique). Des travaux déjà anciens indiquent qu'il existe des variations intonatives significatives qui ne se définissent pas principalement par des variations de fréquence fondamentale ; l'étude de ces phénomènes appelle une prise en compte de la qualité de voix.

Qualité de voix, attitudes et émotions : la qualité de voix est employée, de façon inconsciente ou délibérée, pour véhiculer attitudes et émotions (Fónagy 1983, Fónagy 2001[2] ; voir également le passage en revue des travaux récents par Scherer, Johnstone *et al.* 2003, « Vocal expression of emotion »). Ainsi, Catford note que, dans des langues où la

---

[1] Ainsi, Michel Ferlus postule le développement d'une opposition lexicale de qualité de voix en chinois archaïque (Ferlus 1998a), dans une évolution antécédente à la tonogenèse, dont les deux phases (désormais bien connues) impliquent également l'une et l'autre un stade intermédiaire où la qualité de voix est porteuse de contrastes lexicaux.

[2] Une étude expérimentale récente confirme les observations d'Ivan Fónagy : Gendrot 2003.

voix *murmurée* n'a pas fonction distinctive au plan lexical, elle fonctionne comme un indice intonatif (« In English, and possibly other languages, it is associated with tenderness or mild sexual excitement », Catford 1977:100 ; voir également Fónagy 1983:44-46).

Bilan

Les réflexions qui précèdent visaient à présenter le contexte scientifique de la présente étude, et à rappeler la nécessité d'une prise en compte de la diversité des langues, et de la richesse des phénomènes prosodiques.

4. Notions centrales de la modélisation prosodique

Un bref rappel sur la tradition britannique d'études intonatives vise à souligner les liens privilégiés que celle-ci entretient avec la langue anglaise (son premier et principal objet d'analyse), ce qui suggère l'opportunité, dans une perspective typologique, d'adopter un cadre d'analyse moins spécifique.

4.1. Cadre britannique

Dans la tradition britannique d'études intonatives (dont les ouvrages classiques sont ceux de Palmer 1922, Kingdon 1958, Crystal 1969, O'Connor et Arnold 1973), l'attention se porte (pour schématiser) sur les ordres de faits suivants :

- le découpage en constituants *(phrasing)* ; ces constituants sont appelés unités tonales *(tone units)* par Crystal 1969, groupes de sens *(sense groups)* par O'Connor et Arnold 1973, groupes intonatifs *(intonation groups)* par Cruttenden 1986.
- la détermination des syllabes accentuées et non accentuées de l'énoncé *:* à l'intérieur de chaque groupe intonatif est identifié un noyau *(nucleus)*, « the stressed syllable of the last prominent word in a sense group » (O'Connor et Arnold 1973:271)
- les contours intonatifs par lesquels se réalise l'accent sur le noyau, l'énonciateur ayant le choix à l'intérieur d'un paradigme de mélodies *(tunes)*, aussi appelées contours *(contours)*, ou encore tons *(tones)*, ou parfois accents mélodiques *(pitch accents)*.

Les notions centrales dans l'analyse d'un énoncé sont celles de *pre-head, head,* noyau *(nucleus*, également désigné comme *tonic syllable*) porteur d'une mélodie, et *tail ;* la queue *(tail)* désigne les syllabes inaccentuées qui succèdent au noyau. (De tous les éléments constitutifs de l'unité tonale*, tone unit*, seul le noyau est nécessairement présent.) L'avant-tête *(pre-head)* peut être haute ou basse, la tête basse, rehaussée *(stepping head)* ou glissante *(sliding head)*.

4.2. Cadre de la présente étude

La description qui vient d'être évoquée s'applique particulièrement bien à l'anglais, langue pour laquelle elle a été conçue. D'un point de vue typologique, cette qualité peut se retourner en inconvénient : il nous paraît malaisé d'employer ce modèle à des fins de mise en regard de plusieurs langues, particulièrement s'il s'agit de langues à tons lexicaux. Le cadre retenu pour la présente étude est celui que définissent les travaux de M. Rossi et J. Vaissière (s'inscrivant dans une tradition qui sera présentée plus avant dans la Discussion), qui comme les travaux de l'école britannique d'études intonatives portent principalement sur les langues romanes et germaniques, mais dont les notions centrales nous paraissent plus aisément transposables à d'autres langues. La prosodie telle qu'elle est entendue ici inclut l'*accentuation*, l'*intonation*, et les *facteurs de performance* (dont le rythme et le débit). **L'accentuation est l'ensemble des phénomènes distinctifs lexicaux**. Cette caractérisation exclut l'accent d'insistance (qui ressortit à la dimension *pragmatique* de l'intonation), ainsi que les accents fixes (accents démarcatifs), car ceux-ci n'ont pas fonction distinctive au plan lexical[1]. Inversement, au plan typologique, ce point de vue amène à inclure dans l'accentuation l'ensemble des phénomènes que sont (selon les langues) l'accent lexical, les accents mélodiques, les tons lexicaux ou les registres de qualité de voix. Inclure dans l'*accentuation* les tons lexicaux (qui existent dans des langues comme le naxi et le vietnamien) et les registres de qualité de voix (qui existent dans des langues comme le môn) a quelque chose de paradoxal, *accentuation* évoquant immédiatement *accent*, et non *ton*, encore moins le phénomène peu courant des *registre de qualité de voix*. Ce choix résulte de l'extension par analogie de la catégorie « accentuation » du cadre proposé par M. Rossi et J. Vaissière ; il fait ressortir une caractéristique commune à l'accent lexical (accent libre), aux tons lexicaux et aux registres de qualité de voix, qui est de posséder un statut lexical distinctif. Il pourrait être indiqué de choisir un autre nom pour la catégorie en question ; cela n'a toutefois pas paru avisé au stade actuel du travail, car le choix d'une nouvelle dénomination aurait pu brouiller les correspondances de vocabulaire avec les cadres théoriques choisis pour référence.

L'intonation, domaine d'une grande complexité, inclut les faits de structuration post-lexicale que sont le découpage en constituants (composante *syntaxique* de l'intonation, à fonction *démarcative*) et le marquage de la structure informationnelle (composante *pragmatique* de

[1] La dichotomie établie au plan synchronique entre accent libre (accent lexical) et accent fixe (accent grammatical) ne doit bien sûr pas masquer les liens qu'entretiennent ces catégories au plan diachronique : il est bien connu que l'accent libre peut provenir d'un ancien accent fixe. Cette question est abordée dans la section 3.1.1 de la Discussion (chapitre IV).

l'intonation)[1], ainsi que l'expression des attitudes et émotions (Vaissière 2004 ; Rossi 1999:7, 87 ; Di Cristo 2004:183).

Cette définition correspond à celle donnée par André Martinet (« on classe dans la prosodie tous les faits de parole qui n'entrent pas dans le cadre phonématique », Martinet 1960, § 3.24) ; le tour positif qui lui est donné dans la présente formulation (à la différence de la formulation négative de Martinet) fait d'elle un programme de recherche : au plan phonétique, la prosodie se manifeste par *toutes les variations sous-glottiques, glottiques et supra-glottiques non prévisibles à partir de la connaissance des phonèmes*. **Cette définition programmatique ouvre des perspectives de recherche larges, sans clore l'inventaire des faits prosodiques (ni au plan des fonctions, ni au plan phonétique).** Elle distingue entre faits d'accentuation lexicale et faits intonatifs, et ne préjuge pas de leurs corrélats phonétiques. (Pour un aperçu des nombreuses caractérisations de la prosodie et de l'intonation qui ont été proposées au cours du XX[e] siècle, voir les travaux de Di Cristo et Mario Rossi, qui rappellent notamment les définitions fondatrices des linguistes de l'école de Prague [Rossi 1999:9, 14-55 ; Di Cristo 1975, 2004]).

Les deux paragraphes qui suivent décrivent et commentent les divisions tracées à l'intérieur du domaine intonatif : le découpage en constituants *(phrasing* de l'école britannique) est désigné comme *intonation syntaxique ;* les phénomènes pragmatiques (tels que l'*accent d'insistance* ; le marquage intonatif du statut *focal* d'un constitutant, phénomène très répandu ; et l'*accent contrastif*, typologiquement moins fréquent) constituent l'*intonation pragmatique*.

4.2.1. Intonation syntaxique (= composante syntaxique de l'intonation)

Le terme d'« intonation syntaxique » est employé comme forme brève de *composante morphosyntaxique de l'intonation*. Il désigne le découpage de l'énoncé en constituants sur la base de sa structure morphosyntaxique. Ce terme laisse à désirer dans la mesure où la relation entre frontières morphosyntaxiques et frontières intonatives (en anglais : *intonational boundaries*, ou *intonational junctures*) n'est pas une correspondance bijective, ce sur quoi insistent notamment Philippe Martin, Mario Rossi et Elisabeth Selkirk (Martin 1981, Rossi 1997 ; 1999:132-137, Selkirk 2000:231) ; ainsi, le *mot phonologique* ne correspond pas nécessairement au mot lexical ni à tel ou tel constituant syntaxique. Il serait plus prudent de parler de *découpage en constituants intonatifs* (une expression courante dans les travaux en langue anglaise est *intonational phrasing*). La raison pour laquelle le terme d'*intonation syntaxique* est néanmoins maintenu par M. Rossi (Rossi 1999:131-175) et J. Vaissière est

[1] Par commodité, la composante syntaxique et la composante pragmatique de l'intonation sont respectivement désignées comme *intonation syntaxique* et *intonation pragmatique*. Elles correspondent respectivement à la fonction démarcative et la fonction d'emphase d'Ivan Fónagy (Fónagy 2003, « Des fonctions de l'intonation : essai de synthèse » ; voir également Fónagy 1983, 1989).

que morphosyntaxe et découpage intonatif en constituants sont liés d'une façon régulière, quoique non bijective. Ainsi, en français, la connaissance de la syntaxe d'un énoncé suffit pour synthétiser une courbe de F_0 qui soit jugée acceptable par les auditeurs interrogés (Vaissière 1971).

4.2.2. Intonation pragmatique (= composante pragmatique de l'intonation)

L'intonation pragmatique est présentée ici à la suite de l'intonation syntaxique pour refléter l'intuition selon laquelle la pragmatique vient modifier, parfois au point de les bouleverser, les schémas déterminés par l'intonation syntaxique (Mario Rossi parle d'« intrusion des forces sémantiques »[1]). (Les schémas déterminés par l'interaction de la composante syntaxique et de la composante pragmatique sont eux-mêmes sujets à transformation sous l'effet de l'expression des attitudes et émotions.)

L'intonation pragmatique désigne ce qui dans l'intonation contribue à véhiculer la structure informationnelle de l'énoncé. Ces termes relativement consensuels sont également assez vagues. Les incertitudes de vocabulaire sont une des caractéristiques des études de pragmatique, malgré les trésors d'ingéniosité et de rigueur des jeux de définitions déployés depuis plus d'un siècle (voir notamment Halliday 1967:200, Lambrecht 1994 ; l'histoire des débats sur des notions telles que *thème* et *rhème* remonte à l'antiquité grecque ; voir notamment Prévost 1998). Le souhait d'opérer avec des catégories clairement définies conduit à élaborer un ensemble de définitions rigoureuses ; tel auteur s'appuie sur un jeu de définitions qui lui permet de déterminer à coup sûr, pour les énoncés qu'on lui présente, quel élément a statut de *thème*, de *rhème*, de *focus* ; mais sa réponse sera différente de celle d'un collègue interrogé au sujet des mêmes énoncés. Le linguiste qui tente des découpages ne reposant pas sur des oppositions entre formes linguistiques court le risque de manquer son objet. Certaines langues présentent des catégories linguistiques clairement définies : un exemple en est fourni par le marquage du focus dans la morphologie verbale en wolof (Robert 1986, 2000, Creissels et Robert 1998, Rialland et Robert 2001), par des particules de discours en tamang (Mazaudon 2003) ; d'autres appelleraient un traitement plus nuancé, comme le remarque Marie-Claude Paris au sujet du chinois : « ... plutôt qu'une dichotomie entre le topique et le focus, le chinois présente à mes yeux une échelle de degrés entre différents rôles informationnels » (Paris 1999:202 ; voir également Paris 1997:102-103). La solution adoptée ici consiste à adopter des termes qui paraissent adéquats pour décrire les faits de structure linguistique (dans une perspective typologique), en explicitant l'emploi qui

[1] Le contexte est le suivant : « ... l'étude des interactions entre la syntaxe et la pragmatique a longtemps été négligée. Il nous semble que l'identification des contraintes que chacun de ces deux niveaux fait peser sur l'autre est nécessaire pour évaluer les degrés de liberté de l'intonation, les lois de l'association entre intonation et syntaxe et les modalités d'intrusion des forces sémantiques » (Rossi 1999:9).

en est fait, de façon à ce que le lecteur puisse le cas échéant les retraduire dans un autre cadre de définitions.

Bilan de l'Introduction : Questions centrales

Les principales questions que le volet expérimental doit permettre de traiter sont les suivantes, par ordre croissant d'abstraction à partir des faits expérimentaux :

1) Les locuteurs de langues à tons lexicaux recourent-ils moins que les locuteurs de langues sans tons à la fréquence fondamentale dans la réalisation de l'intonation pragmatique ? La corrélation entre F_0 et intensité est-elle moins forte dans ces langues ? En termes plus généraux, quel type de relation existe-t-il entre spécification lexicale et emploi intonatif ?

2) Quel est le degré de similarité entre les tons lexicaux et les « tons intonatifs » postulés par certaines descriptions ?

3) Comment les diverses caractéristiques des systèmes prosodiques traits prosodiques s'agencent-ils dans une perspective typologique ?

Le chapitre I est consacré à la méthode expérimentale, le chapitre II à un panorama de l'intonation du naxi. Le chapitre III se concentre essentiellement sur l'étude de la réalisation phonétique, dans les trois langues, de mots monosyllabiques ayant statut de *focus* (le *nucleus* de la tradition britannique), cela dans deux contextes d'énonciation qui diffèrent par le degré d'insistance qui porte sur le mot-cible. Une discussion générale est proposée au chapitre IV ; elle s'attache à dégager certaines implications typologiques des observations.

Table des matières du chapitre I (« Méthode expérimentale : enjeux et outils »)

1. Intérêt d'une mesure de la qualité de voix (et de la fréquence fondamentale) par électroglottographie ____ 17

1.1. Difficulté des mesures de la qualité de voix (et secondairement de la fréquence fondamentale) à partir du signal audio ____ 17

1.1.1. Difficulté de l'évaluation de la qualité de voix à partir du signal audio ____ 17

1.1.2. Remarques sur la mesure de F_0 à partir du signal audio ____ 19

1.2. Présentation de l'instrument : principes, histoire, et panorama des méthodes d'analyse du signal électroglottographique ____ 19

1.2.1. Principes et histoire ____ 19

1.2.2. Les méthodes de calcul de coefficients par seuillage du signal électroglottographique présentent une part d'incertitude ____ 20

1.2.3. Quotient ouvert et fréquence fondamentale sont calculables à partir du signal dérivé du signal électroglottographique ____ 22

1.2.3.1. Limites de la mesure, du fait des doubles pics de fermeture et d'ouverture ____ 23

1.2.3.2. Limites de la méthode combinée de calcul du quotient ouvert, dite *de Howard* ____ 25

1.2.3.3. Interprétation du quotient ouvert ____ 26

1.2.4. Réalisation pratique des mesures ____ 29

2. Remarques sur les fréquences formantiques et l'intensité ____ 29

2.1. Les fréquences formantiques et leur mesure ____ 29

2.1.1. Rappels théoriques ____ 29

2.1.2. Difficulté et limites de l'estimation des fréquences de résonance du conduit vocal à partir du signal audio 30

2.1.3. Choix d'une procédure pour l'estimation des fréquences formantiques ____ 33

2.1.3.1. Principes de fonctionnement des outils de détection des formants proposés par le logiciel PRAAT ____ 33

2.1.3.2. Choix d'effectuer une vérification visuelle des résultats ____ 34

2.1.3.3. Emploi d'un second logiciel, pour obtenir un point de comparaison ____ 35

2.2. Mesures d'intensité acoustique globale ____ 36

2.2.1. Variabilité en fonction de la position du locuteur par rapport au micro ____ 36

2.2.2. L'interprétation de la mesure d'intensité ____ 37

Chapitre I. Méthode expérimentale : enjeux et outils

Les expériences réalisées couvrent la fréquence fondamentale, la durée et l'intensité (que l'on peut désigner comme les trois paramètres intonatifs classiques, ceux par exemple que retient Ilse Lehiste dans sa caractérisation des *« Suprasegmentals »* [Lehiste 1970]), mais également la **fréquence des formants** (dont la variation est liée aux déformations des cavités supraglottiques, et par là au changement de timbre des voyelles), et la ***qualité de voix* étudiée par électroglottographie**, technique de mesure de la surface d'accolement des plis vocaux au cours de la phonation. Ce chapitre présente en détail cette dernière méthode (section 1), sur laquelle reposent les mesures de fréquence fondamentale et de qualité de voix, puis les méthodes d'estimation des fréquences formantiques et de l'intensité acoustique (section 2).

1. Intérêt d'une mesure de la qualité de voix[1] (et de la fréquence fondamentale) par électroglottographie

L'électroglottographie permet d'obtenir des indications sur la qualité de voix, et une mesure très précise de F_0. La section 1.1 revient sur le fait qu'il est malaisé (et hasardeux) d'obtenir ces renseignements à partir du signal audio seul.

1.1. Difficulté des mesures de la qualité de voix (et secondairement de la fréquence fondamentale) à partir du signal audio

1.1.1. Difficulté de l'évaluation de la qualité de voix à partir du signal audio

L'approche expérimentale peut consister, soit à rechercher des mesures qui mettent en lumière les corrélats d'oppositions linguistiques établies, soit à mettre au point des outils qui permettent d'analyser des données linguistiques sans avoir de connaissance préalable des catégories en jeu.

La première méthode est relativement simple, mais nécessite que les catégories linguistiques soient définies par avance. Ainsi, Ratree Wayland et Allard Jongman, dans leur étude de l'opposition entre registre de *voix soufflée* et registre de *voix modale* dans un dialecte khmer (Wayland et Jongman 2003), partent de l'opposition catégorielle, établie par des enquêtes non instrumentales, entre voyelles prononcées en voix soufflée et en voix modale ; le travail expérimental consiste à rechercher les corrélats de cette opposition. Les auteurs réalisent sept ensembles de mesures sur le signal audio, qui sont liées de façon plus ou moins directe au débit d'air glottique et à la durée de la phase fermée du cycle glottique :

---

[1] La qualité de voix a été évoquée en Introduction, section 3.4. Rappelons que l'expression est employée ici dans le sens de *mode de vibration des plis vocaux*.

1) amplitude du premier harmonique, H_1, comparée à l'un des paramètres suivants : l'intensité acoustique globale de la voyelle *(RMS amplitude)*, l'amplitude du deuxième harmonique (H_2), ou l'amplitude de l'harmonique le plus saillant dans la région du premier formant (A_1) ou du troisième formant (A_3)
2) rapport entre composante périodique et apériodique (bruit) dans le signal
3) pente spectrale
4) degré de couplage entre trachée et conduit vocal
5) fréquence fondamentale
6) durée
7) intensité acoustique globale

La figure 1.1 (reprise de Wayland et Jongman 2003:19) fournit un exemple de coupe spectrale, qui montre la position et l'amplitude de H_1, H_2, A_1 et A_3.

Pour les données enregistrées par ces auteurs, ce sont les mesures 1 (mesures de pente spectrale : $*H_1 - *H_2$, $*H_1 - *A_1$, $*H_1 - *A_3$) et 7 (mesure de l'intensité) qui sont les plus nettement corrélées à la distinction linguistique qui était établie d'emblée. Dans cette perspective, les catégories linguistiques de départ fournissent la pierre de touche de la fiabilité de l'outil ; les résultats n'éclairent que de façon indirecte les processus phonétiques qui sont en jeu. (La même remarque vaut pour les travaux de Huffman 1987 et Blankenship 2002.)

Le cheminement inverse, qui consiste à caractériser la qualité de voix d'un signal audio sans préjuger de celle-ci, est problématique. La difficulté d'évaluer le degré de souffle *(breathiness)* à partir du signal audio ressort des résultats mitigés obtenus par Shrivastav et Sapienza 2003, qui recourent à deux paramètres : le rapport entre la composante périodique et la composante apériodique du signal, et un paramètre qu'ils baptisent « partial loudness of the periodic signal ». Les incertitudes techniques dans l'estimation de l'amplitude des harmoniques corrigée de l'effet perturbateur du filtrage supraglottique font douter certains chercheurs de la validité de telles mesures (mises en garde notamment formulées par Shinji Maeda et Nathalie Henrich, communication personnelle). Le filtrage inverse n'est guère applicable que pour la voyelle /a/, du fait de la distance qui, dans le cas de cette voyelle, sépare le premier formant des premiers harmoniques. La nasalisation des voyelles nuit également à la précision du filtrage inverse ; or il est difficile de se garantir contre l'éventualité d'une nasalisation, particulièrement dans des langues comme l'anglais, le vietnamien et le naxi, où le trait de nasalité n'est pas pertinent pour la distinction des voyelles (voir les observations de Cohn 1990 sur l'anglais, et ses conclusions générales sur la précision variable du contrôle de la nasalité selon la langue). **S'il s'agit, comme ici, d'étudier de façon quantifiée la qualité de voix de données comprenant des voyelles variées, parfois précédées de consonnes nasales, il est souhaitable d'obtenir un renseignement sur la qualité de voix qui ne présente pas les incertitudes du filtrage inverse et des mesures de**

pente spectrale effectuées sur le signal acoustique. L'électroglottographie présente un intérêt dans cette perspective.

1.1.2. Remarques sur la mesure de F_0 à partir du signal audio

La détection de F_0 à partir du signal audio ne va pas de soi, du fait du filtrage du signal de source par les cavités supraglottiques, qui transforme l'impulsion glottique, signal relativement simple, en un signal complexe. Une mesure de F_0 est proposée par les nombreux logiciels d'analyse du signal acoustique ; parmi les plus répandus, citons PRAAT, WINPITCH, SNOORI, UNICE, XWAVES, SOUNDSWELL, EMU, SPEECH FILING SYSTEM. Les réussites techniques ne doivent pas faire oublier que l'estimation de la durée des cycles glottiques à partir du signal audio peut poser problème, en particulier dans les cas de phonation non modale. Une solution assez répandue dans les centres de recherches phonétiques avant l'ère des logiciels d'analyse du signal consistait à placer un micro sur la gorge du locuteur à la hauteur du larynx (micro appelé *accéléromètre*), ce qui fournit un signal d'interprétation plus aisée que le signal audio, puisque le bourdonnement produit par la vibration des plis vocaux est enregistré au plus près de la source. L'électroglottographie offre pareillement une mesure de fréquence fondamentale en amont du filtrage par les cavités supraglottiques. **Sur le signal électroglottographique, chaque cycle glottique se détache nettement, du début du voisement jusqu'au dernier cycle glottique** (voir par exemple, au chapitre III, les figures 3.61 et 3.62a, qui présentent le signal électroglottographique en regard du signal audio correspondant), et ce même dans les cas de glottalisation. (Signalons que les enregistrements électroglottographiques sont parfois employés comme pierre de touche dans l'évaluation des algorithmes de détection à partir du signal audio.)

1.2. Présentation de l'instrument : principes, histoire, et panorama des méthodes d'analyse du signal électroglottographique

Notre méthode d'analyse du signal électroglottographique repose sur la détection des instants de fermeture et d'ouverture de la glotte, à l'aide de la *dérivée du signal électroglottographique*. Elle est exposée en **section 1.2.3, seule section du chapitre I dont la lecture soit un préalable indispensable à la bonne compréhension de la suite de la thèse**. Les sections 1.2.1 et 1.2.2 présentent les questions auxquelles la méthode d'analyse retenue apporte une réponse, de façon à mettre en lumière son originalité et son intérêt.

1.2.1. Principes et histoire

L'électroglottographe, inventé par P. Fabre (Fabre 1957), est un appareil d'utilisation aisée, qui se laisse assez facilement oublier au cours d'une séance d'enregistrement. Un courant de haute fréquence et de très faible intensité circule entre des électrodes placées de part et d'autre du larynx (les photographies 1 à 6 p. 436 illustrent le positionnement des électrodes).

Plus la surface de contact entre les plis vocaux est grande, plus la résistance au courant est faible. Les figures 1.2 et 1.3 (reprises de Henrich 2001:94-95) mettent en correspondance des images du larynx par cinématographie ultra-rapide avec un signal électroglottographique enregistré simultanément (ces figures représentent également la *dérivée* de ce signal, dont le mode de calcul et l'utilité pour l'analyse sont présentés ci-après, section 1.2.3).

Le signal électroglottographique est enregistré simultanément avec le signal audio. Il peut être visualisé pendant l'enregistrement avec un oscilloscope, et par la suite avec les mêmes outils que le signal audio. L'électroglottographe est aussi appelé *laryngographe* : Adrian Fourcin souligne en effet que l'appareil mesure une impédance électrique à travers le larynx dans son ensemble (structure complexe, dont les plis vocaux ne sont qu'une partie ; voir la figure 1.4), d'où son choix du terme *laryngographe* (Fourcin 1971). C'est un appareil sans danger ; les électrodes peuvent donc rester appliquées sur le cou tout au long d'une séance d'enregistrement. Le modèle employé est un électroglottographe à deux canaux (EG2 ; voir Rothenberg 1992) : il comporte deux paires d'électrodes de façon à limiter l'incidence des mouvements verticaux du larynx sur le signal obtenu.

Plusieurs études établissent que le signal électroglottographique varie en fonction de la surface d'accolement des plis vocaux. Il renseigne donc de façon dynamique sur la variation de la surface d'accolement des plis vocaux. Au sujet des informations apportées par ce signal, diverses réponses sont apportées par les travaux récents (dont Gilbert, Potter *et al.* 1984, Titze 1990 et Baken 1992). **Les paramètres relevés sur le signal électroglottographique demeurent très variés.** Pour faire ressortir l'originalité et l'intérêt de la méthode adoptée ici, il est utile de passer en revue les principales autres méthodes proposées à ce jour.

1.2.2. Les méthodes de calcul de coefficients par seuillage du signal électroglottographique présentent une part d'incertitude

Le principe de ces méthodes consiste à faire passer une ligne horizontale à une certaine hauteur relativement aux minima et maxima atteints par le signal électroglottographique, et à relever les points d'intersection entre cette ligne et le signal. La figure 1.5 en fournit une illustration. La réflexion sur une telle mesure par seuillage sera menée ici en prenant l'exemple de la méthode employée par John Esling dans un travail visant à caractériser la forme des signaux électroglottographiques correspondant à divers modes phonatoires (Esling 1984). L'auteur entreprend de mesurer, pour chaque période, le « rapport entre la durée de montée et la durée de descente » (en anglais: « ratio of rise-time to fall-time (RT/FT) »). La première de ces valeurs est la durée qui sépare le point le plus bas du point le plus haut sur le signal électroglottographique, la seconde la durée qui sépare le point le plus haut du point le plus bas suivant (RT = « duration from the low point to the high point of the signal », FT = « duration from the high point to the low point »). Le rapport de ces deux valeurs, RT/FT, fournit une indication sur la rapidité relative des phases de fermeture et d'ouverture de la glotte. La variabilité constatée sur le signal électroglottographique aux environs du

point le plus haut et du point le plus bas conduit J. Esling à effectuer deux corrections préliminaires avant application de la méthode de seuillage (illustrées sur la figure 1.5) :

> 1) a base-line is drawn to transect the point in each period at which the rapid rise begins, in order to eliminate the base-line variation that takes place during the open phase of the cycle.
> 2) only 80% of the signal is measured, from 10% above base-line to 10% below signal peak, with a view to correcting for possible artefacts near the extremes (Esling 1984:61)

En dernière analyse, le choix empirique du placement du seuil représente une approximation. Comme l'explique l'auteur, le paramètre RT *(Rise-Time)* est seulement une approximation assez grossière de la durée de la phase de fermeture de la glotte, de même que FT *(Fall-Time)* pour l'ouverture. Le signal électroglottographique de la figure 1.1 montre la limite de cette méthode : la partie du signal annotée du signe (F) (pour *Fall*) sur la figure possède la pente négative la plus forte de tout le cycle glottique, et correspond donc à l'intervalle temporel où les plis vocaux se séparent avec la célérité la plus élevée (aussi appelé « instant d'ouverture » dans les publications spécialisées) ; or cette partie du signal se trouve au-dehors du cadre obtenu en appliquant la méthode de seuillage, et n'est donc pas prise en compte[1]. Ce problème est de taille, du fait de la forte variabilité de l'allure des signaux électroglottographiques (variabilité entre locuteurs et entre qualités de voix).

La mesure résultante fait ressortir une différence entre les types de phonation que l'auteur souhaitait caractériser. Néanmoins, il s'agit là des extrêmes de la qualité de voix que peut réaliser le larynx humain (voix extrêmement *soufflée*, extrêmement *craquée*, etc.). **Dans l'exploration des variations de qualité de voix présentes dans la parole ordinaire, il paraît utile de disposer d'un outil de mesure assez fin pour situer un échantillon de parole en un point précis de l'échelle qui sépare la voix modale (« normale ») des types phonétiques extrêmes que sont la voix *tendue*, la voix *soufflée* ou la voix *laryngalisée*.** Dans les langues où la qualité de voix joue un rôle linguistique (typiquement, un rôle distinctif au plan lexical), les locuteurs ne la réalisent pas systématiquement de façon canonique (pas plus que les phonèmes ne sont systématiquement réalisés de façon canonique). Resurgit alors le problème évoqué à propos des mesures à partir du seul signal audio : la mesure de quotient ouvert par seuillage permet de distinguer (*modulo* des incertitudes expérimentales relativement élevées) certains types phonétiques extrêmes, mais elle paraît en revanche fragile comme méthode exploratoire dans une recherche qui nécessite une estimation fine de

[1] Précisons que le signal représenté sur la figure n'a pas été choisi à dessein pour mettre en défaut la méthode, mais s'est trouvé être le premier auquel nous l'avons appliquée. Le choix de la méthode de John Esling pour illustrer les limites des méthodes de seuillage appliquées à l'électroglottographie ne vise nullement à porter une critique *ad hominem :* il se trouve qu'ayant envisagé de suivre la méthode qu'il décrit, nous avons rencontré les difficultés rapportées ici, qui sont communes aux méthodes par seuillage.

la qualité de voix. Ainsi, l'étude de la qualité de voix de syllabes anglaises ou naxi menée au chapitre III vise à déceler un éventuel emploi prosodique des nuances de la qualité de voix, sans préjuger du caractère catégoriel des faits observés, ni des désignations à adopter pour décrire les éventuelles différences. En outre, la signification physiologique des coefficients calculés dans la méthode par seuil évoquée ci-dessus est loin d'être transparente. Les mêmes commentaires s'appliquent aux autres méthodes par seuillage, telles que celles appliquées par Justin Watkins à l'étude de langues à registre de voix (Watkins 1999, 2002).

Au vu de ces incertitudes, des spécialistes (en particulier dans le domaine médical) mettent en garde contre les interprétations erronées du signal électroglottographique, qui ne renseigne qu'indirectement sur le flux d'air qui passe à la glotte[1]. La section 1.2.3 souligne que certains paramètres glottiques peuvent pourtant être calculés de façon précise à partir du signal électroglottographique ; telle est la méthode qui a été retenue ici.

1.2.3. Quotient ouvert et fréquence fondamentale sont calculables à partir du signal dérivé *du signal électroglottographique*

Certains paramètres glottiques peuvent être calculés de façon précise à partir du signal électroglottographique, par l'utilisation de son signal *dérivé* (voir Henrich, d'Alessandro *et al.* 2004, et références citées). La dérivée du signal est obtenue en calculant la différence entre les échantillons successifs, ce qui revient à calculer la pente du signal électroglottographique en chacun de ses points. Cette dérivée présente généralement deux pics par période, l'un positif, l'autre négatif (voir figure 1.6). **Des expériences combinant électroglottographie et visualisation des plis vocaux au cours de la phonation ont permis de montrer que l'un des pics de la dérivée du signal électroglottographique correspond à la phase la plus rapide de la fermeture de la glotte (« pic de fermeture »), tandis que l'autre correspond à la phase la plus rapide de l'ouverture de la glotte (« pic d'ouverture »)**, comme le montrent les figures 1.2 et 1.3. La figure 1.7 montre un exemple de signal dérivé comportant des pics bien marqués. Cette figure montre également qu'un cycle glottique peut être divisé en une *phase fermée* et une *phase ouverte* sur la base de la détection des pics d'ouverture et de fermeture. **Cette division permet le calcul d'un paramètre dont l'importance acoustique est bien établie : le quotient ouvert** (dont l'interprétation est discutée ci-dessous, section 1.2.3.3), **égal à la durée de la phase ouverte (entre un pic d'ouverture et le pic de fermeture suivant) divisée par la longueur totale de la période (définie comme la durée séparant deux pics de fermeture successifs).**

[1] Voir en particulier Colton et Conture 1990 (« Problems and Pitfalls of Electroglottography »). Robert Orlikoff déconseille l'emploi de l'électroglottographe dans l'établissement de diagnostics cliniques (Orlikoff 1998, « Scrambled EGG: The Uses and Abuses of Electroglottography »). Presque cinquante ans après la mise au point du premier électroglottographe, un travail récent (Chen, Robb *et al.* 2002) se limite encore au relevé manuel d'un seul paramètre, qu'il est difficile de relier à une réalité physiologique : la distance entre le point le plus haut et le plus bas au cours d'une période du signal électroglottographique.

> Quand la glotte est fermée, le débit est soit nul, soit constant (cas d'une fuite d'air permanente au niveau des aryténoïdes). L'instant d'ouverture glottique est défini par l'instant d'augmentation du débit glottique par rapport à sa valeur minimale. Physiologiquement, cela correspond à l'instant où les cordes vocales commencent à se séparer sur leur partie supérieure. L'instant de fermeture glottique est associé au maximum de l'excitation glottique, c'est-à-dire l'instant où la dérivée du débit glottique atteint son minimum. (Henrich 2001:26)

La *phase ouverte* et la *phase fermée* ne doivent pas être confondues avec la phase de fermeture et la phase d'ouverture. La phase d'ouverture est définie entre l'instant d'ouverture glottique et l'instant du maximum de débit. La phase de fermeture est définie entre l'instant du maximum de débit et l'instant de fermeture glottique. Ces deux phases ne sont pas discutées dans la présente étude car leur estimation nécessiterait de connaître le débit glottique, ce que ne permet pas l'électroglottographie.

Il importe d'indiquer comment sont traités les cas dans lesquels le signal électroglottographique et sa dérivée ne présentent pas la même netteté que l'exemple présenté en figure 1.6.

1.2.3.1. Limites de la mesure, du fait des doubles pics de fermeture et d'ouverture

Dans certains cas, les pics sur la dérivée du signal électroglottographique sont peu saillants, ou dédoublés. Cela n'est pas un artefact, mais tient à la façon dont les plis vocaux s'accolent et se séparent, une assez grande variété s'observant quant à la partie de la glotte qui apparaît en premier à l'ouverture, et qui disparaît en premier à la fermeture (voir Henrich 2001 ; Henrich, Gendrot *et al.* 2004). Le choix de Nathalie Henrich, qui étudie essentiellement la voix chantée, consiste à ne retenir, dans les mesures, que les périodes pour lesquelles les pics d'ouverture comme les pics de fermeture sont uniques et bien marqués. En voix parlée, la vibration des plis vocaux est moins régulière, et le dédoublement (voire triplement) des pics de fermeture paraît plus fréquent (le phénomène varie d'un locuteur à l'autre). La figure 1.7 montre le signal dérivé correspondant à une rime syllabique au ton montant (produite par le locuteur naxi M4) : l'amplitude des pics de fermeture est très basse en milieu de rime. La figure 1.8 montre le signal électroglottographique des huit périodes au cours desquelles la fermeture est la moins bien marquée : il apparaît que chaque cycle glottique se détache nettement ; il n'y a nullement interruption du voisement, de sorte qu'il ne paraît pas légitime d'exclure cet item des résultats (cela d'autant plus que le phénomène est très fréquent dans la production du ton lexical montant chez le locuteur concerné).

Dans la présente étude, le niveau d'exigence a donc été placé plus bas que dans l'étude de la voix chantée menée par Nathalie Henrich *(op. cit.)* : les pics d'ouverture doubles ou imprécis ont été éliminés, mais les pics de fermeture doubles ou triples n'ont pas été exclus des calculs. En pratique, l'utilisateur du programme réalisé pour la présente étude (<peakdet>, décrit en section 4.1 de l'Annexe 2), a le choix entre les méthodes de calcul suivantes :

- retenir le pic le plus élevé
- retenir le premier pic
- retenir le dernier pic
- retenir le barycentre des pics détectés.

La dernière de ces méthodes appelle explication. La figure 1.9 propose un exemple. Sur cette figure, le décompte des cycles glottiques sur le signal dérivé du signal électroglottographique ne pose pas de problème : sur la portion de signal représentée, quatre pics positifs dépassent au-dessus de la ligne rouge qui matérialise le seuil employé pour leur détection. Le détail de la forme de ces quatre pics montre qu'ils sont dédoublés. Dans la méthode des barycentres, comme dans les autres méthodes, les pics sont d'abord détectés (par un seuil qui passe à ¼ du maximum du signal dérivé pour l'ensemble de l'item considéré—typiquement, toute une rime syllabique). Ensuite, pour chacun des pics détectés, le programme recherche son maximum (point le plus haut), puis effectue un second seuillage, à 70% de l'amplitude du point le plus haut. Ce second seuillage est matérialisé en vert sur la figure, pour chacun des pics. Il aboutit à la détection de n dentelures qui dépassent le seuil ; pour chacune de ces dentelures, le maximum local est relevé (amplitude et position). Par exemple, le premier pic de fermeture comporte deux dentelures (deux « petits pics ») ; leur hauteur relative est indiquée sur la figure sous forme de pourcentage. Une fois reconnue la présence de ces deux pics lors de la première fermeture, leur moyenne pondérée est calculée ; cela aboutit à une valeur temporelle (et une valeur d'amplitude). Dans le cas des trois dernières fermetures, les dentelures ont une amplitude inférieure à 70% de celle du maximum local, de sorte que ce dernier est seul retenu[1].

Cette méthode a pour effet un lissage des valeurs de fréquence fondamentale dans les cas où un pic secondaire apparaît et gagne en amplitude d'un cycle à l'autre jusqu'à devenir le pic principal. Ce phénomène, relativement fréquent, est illustré par la figure 1.10 (à ce sujet, voir Henrich, Gendrot *et al.* 2004, « Characterization of features observed on the derivative of EGG signal by the use of high speed cinematography »).

Cette approximation dans la détection des fermetures glottiques représente une limite de la méthode par détection des pics sur le signal électroglottographique ; elle paraît justifiée pour le type de données étudié ici, dans la mesure où une extrême précision dans la mesure de F_0 n'est pas essentielle pour faire ressortir les faits linguistiquement significatifs. (Comme le remarque Wolfgang Hess au terme d'un passage en revue approfondi des méthodes de détection de la fréquence fondamentale, le choix d'un algorithme de détection doit se faire au vu des données et de l'objectif de l'étude : il n'existe pas de méthode unique qui soit en

[1] D'autres choix seraient possibles, par exemple la prise en compte de la forme de chacune des dentelures (en relevant tous les changements de signe de la dérivée seconde du signal électroglottographique), mais les conséquences sur les valeurs résultantes seraient minimes, et aucune raison de principe n'encourage à faire ce choix. (Le fonctionnement du programme que nous avons employé est décrit plus en détail dans l'Annexe 2.)

elle-même supérieure aux autres [Hess 1983]). Le caractère semi-automatique de la mesure permet à l'utilisateur de modifier son choix de méthode de détection des pics de fermeture s'il constate des valeurs de F_0 en dents de scie. Si l'on se fixait pour objectif une mesure de F_0 d'une précision accrue, il serait envisageable d'employer un algorithme de détection par corrélation (tel que celui décrit par Henrich, d'Alessandro *et al.* 2004) pour les passages de vibration quasi-périodique des plis vocaux, et de recourir à l'algorithme par détection des pics employé ici (décrit en section 4.1 de l'Annexe 2) pour le traitement des autres cas : passages de voisement dont le coefficient de corrélation[1] est faible (inférieur à 0.5), et premiers et derniers cycles glottiques.

1.2.3.2. Limites de la méthode combinée de calcul du quotient ouvert, dite *de Howard*

La méthode dite de Howard, qui paraissait très prometteuse, n'a pas fourni les résultats escomptés. Il s'agit d'une méthode qui devait permettre d'obtenir une approximation de l'instant d'ouverture même dans les cas où la méthode de détection de l'ouverture sur la dérivée du signal électroglottographique ne s'applique pas du fait de l'absence de pic d'ouverture bien marqué ; cette méthode est couramment désignée comme *méthode de Howard*, D.M. Howard ayant utilisé et approfondi cette méthode inventée par P. Davies et des collaborateurs (voir Davies, Lindsey *et al.* 1986, Howard, Lindsey *et al.* 1990 et Howard 1995). Elle consiste à détecter les fermetures sur le signal dérivé, et les ouvertures par une méthode de seuil sur le signal électroglottographique (non dérivé). Vu-Ngoc Tuân a créé un programme implémentant une variante de cette méthode, dans laquelle la droite de seuillage est une parallèle à la ligne de base tracée d'une période à l'autre en passant par le point auquel a lieu la fermeture glottique (point sur le signal électroglottographique correspondant au pic positif sur le signal dérivé ; voir Vu-Ngoc, d'Alessandro *et al.* 2005). Au bilan, les résultats ont été relativement décevants du fait de l'influence considérable qu'a l'amplitude globale du signal électroglottographique sur les valeurs obtenues : en début et en fin de voisement, en particulier, la différence d'amplitude d'une période à l'autre aboutit à des valeurs de quotient ouvert manifestement erronées. Plus fondamentalement, cette méthode n'échappe pas à l'incertitude qui pèse sur toute méthode par seuillage, pour les raisons exposées au paragraphe 1.2.2 ; un intérêt de la méthode de Howard est qu'elle permet de calculer une estimation du quotient ouvert même dans les cas où le pic d'ouverture n'est pas décelable, mais l'absence de pic d'ouverture bien marqué est précisément un indice du fait qu'il n'est pas possible d'opposer nettement une phase ouverte et une phase fermée pour le cycle glottique concerné, et qu'il est donc prudent de renoncer à calculer le quotient ouvert pour ce cycle.

---

[1] Au sujet du choix de ce seuil, voir la description de l'algorithme DECOM par Henrich, d'Alessandro *et al.* 2004.

1.2.3.3. Interprétation du quotient ouvert

Gardant en mémoire qu'il n'y a pas de lien direct et univoque entre tel ou tel paramètre, le système musculaire de la phonation et l'effet produit, il est néanmoins possible d'interpréter le quotient ouvert comme indice du degré d'abduction des plis vocaux (Rothenberg et Mahshie 1988, « Monitoring vocal fold abduction through vocal fold contact area »). Toutes choses égales par ailleurs, une augmentation de quotient ouvert signale un *relâchement* (la phonation se rapprochant de la *voix soufflée*, à fort débit d'air) une diminution de quotient ouvert un *serrage* de la voix (la phonation se rapprochant de l'extrême que représente la constriction glottale).

L'interprétation des valeurs de quotient ouvert demande certaines précautions. Une même valeur de quotient ouvert n'appelle pas la même interprétation selon qu'elle est relevée chez une femme ou chez un homme, dans le haut ou le bas de la tessiture du locuteur. Un exemple en est fourni par les réalisations de la phrase chinoise (1) par une locutrice du mandarin standard (expérience-pilote réalisée au Laboratoire d'Acoustique Musicale, combinant enregistrement audio et enregistrement électroglottographique) :

(1)	ta[1]	yao[4]	na[4]	zhi[1]	bi[3].	(他要那只笔。)
	3[e] sg.	vouloir	déictique	classificateur	stylo, pinceau	

Selon qu'une focalisation contrastive porte sur l'une ou l'autre des syllabes (le pronom sujet, le verbe, le déictique, ou le nom), le sens de la phrase est : (a) C'est lui (qui veut ce stylo). / (b) (Mais oui,) il le veut, ce stylo ! / (c) Il veut ce stylo-ci (et pas un autre). / (d) Il veut ce stylo (pas un autre objet)[1].

La figure 1.11 représente les courbes de F_0 et de DECPA[2] correspondant à une réalisation de l'interprétation (d), avec pour contexte d'élicitation la question « Qu'est-ce que c'est qu'il veut ? c'est ce stylo ou cette trousse ? »[3]. (L'enregistrement audio correspondant est

---

[1] L'affaire se complique quelque peu du fait que la focalisation contrastive en position finale est susceptible de plusieurs interprétations, ce qui rejoint les observations souvent citées de Jackendoff 1972 sur l'anglais : « He was warned to look out for an ex-convict with a RED shirt » attire l'attention sur la couleur rouge ; en revanche, si c'est le nom plutôt que l'adjectif qui est mis en valeur (« He was warned to look out for an ex-convict with a red SHIRT »), plusieurs interprétations sont possibles : dans les termes de Cruttenden 1986, il peut s'agir d'un focus étroit, *narrow focus,* aussi bien que d'un focus large, *broad focus* (à ce sujet, voir également Rossi 1999:111). Précisons en outre que l'étiquette « focalisation contrastive » employée pour décrire l'élément qui est ici mis en valeur au plan intonatif est discutable ; ce n'est pas ici le lieu d'entrer dans la discussion des divers cadres théoriques (dont celui de Lambrecht 1994 est l'un des plus aboutis), dans la mesure où cette discussion est sans incidence directe sur la question traitée ici.

[2] DECPA est un paramètre calculé à partir de l'électroglottographie : il s'agit de l'amplitude des pics positifs sur le signal dérivé ; ce paramètre est décrit plus en détail dans l'Annexe 2, section 3.

[3] Les figures 1.11 et 1.12 sont reprises d'une communication de colloque, « A Measurement from Electroglottography: DECPA, and its Application in Prosody » (Michaud 2004b) ; par commodité, elles n'ont pas été retraduites de l'anglais.

1d_A.wav, l'enregistrement électroglottographique 1d_E.wav.) DECPA reflète la *vitesse maximale d'acroissement de la surface d'accolement des plis vocaux lors de la fermeture glottique*. Ce paramètre est moyennement élevé sur les quatre premières syllabes, qui sont réalisées en *voix de tête* ; sur la dernière syllabe, DECPA est élevé en début et en fin (portions réalisées en *voix de poitrine*), très bas au cœur de la syllabe (portion réalisée en *voix craquée*). La figure 1.12 montre comment, en réponse à la question « Il veut lequel : celui-ci ou celui-là ? » (interprétation c), les trois premières syllabes, réalisées en voix de tête, ont des valeurs moyennes de DECPA, la quatrième, réalisée en voix de poitrine, des valeurs très élevées, tandis que la quatrième, en voix craquée, présente des valeurs basses de bout en bout. (L'enregistrement audio correspondant est 1c_A.wav, l'enregistrement électroglottographique 1c_E.wav.) Ces figures illustrent l'ampleur des variations du signal électroglottographique selon que la locutrice parle en voix de tête, voix de poitrine ou voix craquée. Les observations sur les données de locutrices de langue naxi (locutrices F1 et F2) confirment le passage fréquent d'un mécanisme laryngien à l'autre.

Les termes « voix de tête » et « voix de poitrine » employés au sujet d'un locuteur/chanteur et d'une locutrice/chanteuse n'ont pas la même signification : ce que l'on nomme « voix de poitrine » et « voix de tête » chez l'homme correspond simplement à deux registres résonantiels distincts, tous deux produits avec le même mécanisme laryngé. Les termes « voix de tête » et « voix de poitrine » peuvent donc donner lieu à des malentendus. Ils s'inscrivent, dans l'emploi qui en est traditionnellement fait dans le domaine de la voix chantée, à l'intérieur d'un paradigme qui comporte également d'autres catégories, telles que la « voix mixte », laquelle ne correspond en fait pas à un mécanisme laryngien distinct, mais est le fruit d'un entraînement des chanteurs (voir Castellengo, Chuberre *et al.* 2004). Il paraît utile, afin de prévenir les confusions que peut engendrer l'emploi de termes comme « voix de poitrine » et « voix de tête », de présenter la classification proposée par B. Roubeau et M. Castellengo, qui s'appuie sur des travaux antérieurs (en particulier Hirano, Vennard *et al.* 1970, Titze 1994 ; voir une revue bibliographique dans Roubeau 1993) et d'études électroglottographiques de la voix chantée (Roubeau, Chevrie-Muller *et al.* 1987, 1991). Cette classification distingue entre *mécanismes phonatoires* (ou *mécanismes laryngés*) :

- le mécanisme 0 ou voix craquée *(creaky voice)*[1]. (Nous désignons le passage en mécanisme 0 comme **laryngalisation**.)
- le mécanisme 1 correspond à ce qui est communément nommé « voix de poitrine » chez la femme ; il est employé par les locutrices et locuteurs dans le bas de la tessiture.
- le mécanisme 2 coïncide pour l'essentiel avec la « voix de tête » chez la femme ; il est employé par les locutrices et locuteurs dans le haut de la tessiture[1].

[1] Il existe également un mécanisme 3 ou « voix de sifflet », employé dans la partie la plus élevée (suraiguë) de la tessiture. Le mécanisme 3 n'est pas abordé ici, n'étant pas représenté dans notre corpus.

A ce jour, l'usage de ce mode de classification ne s'est pas imposé dans la communauté des chercheurs ; il est néanmoins employé ici, car il nous paraît aussi éclairant pour la voix parlée que pour la voix chantée, et peut assez commodément être mis en rapport avec les données électroglottographiques.

En voix craquée (mécanisme 0), le calcul du quotient ouvert pose problème, dans la mesure où le modèle d'une fermeture unique et d'une ouverture unique par cycle glottique ne s'applique pas. Les phénomènes de changement de mode vibratoire des plis vocaux sont depuis longtemps l'objet d'études (Catford 1964, 1977:95-106 ; Laver 1980; Titze 1994, 1995 ; Ladefoged et Maddieson 1996:73). La voix craquée est parfois appelée *friture vocale (vocal fry)* (Hollien, Moore *et al.* 1966) ou *laryngalisation*. Une observation détaillée de ce mode vibratoire, sur la base d'images de la glotte par cinématographie ultra-rapide, est proposée par Gendrot, Henrich *et al.* 2004. Les signaux électroglottographiques de ce mécanisme laryngien sont très variés ; Philippe Martin (communication personnelle) y voit l'image d'un *chaos*, dans lequel les événements (passage ou non d'une bulle d'air, fermeture plus ou moins complète, temps de fermeture plus ou moins long) sont pour une part aléatoires. Dans le présent travail, aucune estimation de quotient ouvert n'a été tentée sur la voix craquée[1].

En mécanisme 1, la plage de variation de quotient ouvert observée en parole s'étend de 35% (voix très *pressée*, qui se rapproche d'une constriction glottale) à 70%, voix très *détendue* et à débit d'air élevé (Titze 1988:236). Nos données comportent un grand nombre d'exemples dans lesquelles des valeurs plus extrêmes sont observées sur les premiers ou les derniers cycles glottiques : le premier cycle après une consonne aspirée peut comporter une phase fermée très courte pour une phase ouverte très longue, le rapport étant alors supérieur à 70% ; de même pour les deux ou trois dernières périodes de voisement avant que le voisement ne laisse la place à un bruit de friction. A l'inverse, une constriction glottale en fin de syllabe peut aboutir à une diminution rapide des valeurs de quotient ouvert, jusqu'à des valeurs inférieures à 20%. Ces valeurs extrêmes ne s'observent qu'aux marges des intervalles voisés : lors du début et de l'interruption du voisement, non sur des intervalles de voisement stable[2].

---

[1] Il paraît essentiel, pour l'étude des phénomènes linguistiques, d'opposer la voix craquée à la *constriction glottale*, forme extrême de tension : la voix craquée en elle-même ne représente pas un mode phonatoire *tendu* ; elle peut au contraire correspondre à une forme de détente (Ivan Fónagy la rapproche du ronronnement du chat [Fónagy 1983:44]). L'opposition entre constriction glottale et voix craquée sera présentée à l'occasion de l'exposition des résultats expérimentaux sur la langue vietnamienne : chez les jeunes locuteurs masculins de Hanoi, ces deux qualités de voix sont associées à deux tons lexicaux différents (voir Michaud 2004a :138-139, Vu-Ngoc Tuân, d'Alessandro *et al.* 2005, et ci-dessous chapitre III, section 3.3.1 et figure 3.61bis).

[2] La notion de « voisement stable » est assurément très relative : le voisement n'est jamais parfaitement périodique, et il est particulièrement variable en voix parlée. Il paraît indiqué de choisir pour critère la

En mécanisme 2 (« voix de tête » pour les locutrices ; il n'existe pas de terme précis et largement accepté pour la phonation en mécanisme 2 des hommes)**, le quotient ouvert est globalement plus haut et moins variable.** Il est donc essentiel, dans la comparaison de valeurs de quotient ouvert, de tenir compte du mécanisme laryngien employé. La transition entre mécanismes est fréquente, non seulement chez les locutrices, mais aussi chez les locuteurs ; elle doit être prise en compte dans la comparaison des valeurs de quotient ouvert.

1.2.4. Réalisation pratique des mesures

Au terme de l'exposé sur l'emploi du signal électroglottographique, quelques précisions au sujet de la réalisation pratique des mesures peuvent être utiles.

Les bornes de début et de fin des intervalles à analyser sont posées au vu du signal acoustique (dans le logiciel SoundForge) et avec l'aide de l'écoute (et secondairement de spectrogrammes).

Les outils logiciels élaborés pour l'analyse du signal électroglottographique (en collaboration avec d'autres chercheurs) **sont disponibles sur un site dédié à l'analyse du signal électroglottographique**[1]. **Leurs principes de fonctionnements sont détaillés en section 4.1 de l'Annexe 2**.

2. Remarques sur les fréquences formantiques et l'intensité

2.1. Les fréquences formantiques et leur mesure

2.1.1. Rappels théoriques

Les formants des voyelles sont les régions où les harmoniques sont renforcés du fait des résonances du conduit vocal ; la figure 1.13, reprise de Fant 1960:19, rappelle le modèle source-filtre de la parole. Les formants renseignent sur la configuration du conduit vocal[2]. Néanmoins, les relations entre formants et configurations du conduit vocal ne sont pas biunivoques (voir en particulier Atal, Chang *et al.* 1978), d'où la nécessité d'une grande prudence dans l'interprétation en termes articulatoires des fréquences formantiques relevées.

---

possibilité de maintenir un certain type de phonation pendant une seconde ou plus ; c'est dans ce cadre qu'il faut entendre l'estimation proposée par Titze d'un quotient ouvert variant de 35% à 70%.

[1] http://www.lam.jussieu.fr/src/Membres/Henrich/egg/, site créé en collaboration avec Nathalie Henrich, Vu-Ngoc Tuân et Cédric Gendrot.

[2] L'ouvrage central des études reliant articulation et acoustique est le livre de Gunnar Fant, *Acoustic theory of speech production, with calculations based on X-ray studies of Russian articulations* (Fant 1960). La réflexion de Pierre Delattre (par exemple Delattre 1958), nourrie par des expériences en synthèse de la parole, est également fondatrice. Plusieurs manuels présentent les notions essentielles d'acoustique phonétique : le livre de Peter Ladefoged *Elements of Acoustic Phonetics* (Ladefoged 1962) ; les manuels de Ladefoged 1975 et Kent et Read 1992.

Il est nécessaire de garder à l'esprit que **la fréquence des formants ne fournit pas d'indications directes sur le mouvement des articulateurs** ; ainsi, le premier formant renseigne sur l'aperture de la voyelle, qui est déterminée par le degré d'aperture de la mâchoire mais aussi par le degré d'abaissement de la langue. Pour les voyelles antérieures fermées, telles que /i/, le premier formant (ci-après F_1) est dû à une résonance de type Helmholtz, et dépend donc de la configuration des lèvres ; il n'est donc pas surprenant que l'évolution du premier formant puisse aller en sens contraire de l'aperture de la mâchoire dans le cas de voyelles fermées. En position prosodique forte (typiquement, sous accent d'insistance), une aperture accrue de la mâchoire a été observée *pour toutes les voyelles* (à divers degrés, l'accroissement d'aperture étant plus fort pour les voyelles ouvertes) ; celle-ci va de pair avec une réalisation de la cible vocalique dans laquelle le mouvement de la langue est d'une ampleur accrue, de sorte que le premier formant de cette voyelle n'est pas nécessairement plus élevé sous accent d'insistance (Erickson, Fujimura *et al.* 2002 ; Erickson, Iwata *et al.* 2004 ; Cho 2005).

Il importe en outre de rappeler certaines limites de l'estimation des fréquences de résonance du conduit vocal à partir du signal acoustique.

2.1.2. Difficulté et limites de l'estimation des fréquences de résonance du conduit vocal à partir du signal audio[1]

La détection de la fréquence des formants à partir du signal acoustique peut paraître une opération de routine : divers logiciels proposent une détection automatique de formants à partir du signal acoustique. Les résultats peuvent paraître d'une extrême précision : le logiciel PRAAT fournit des valeurs à une précision de plus de dix décimales (par exemple : F_1 = 543.8785254660551 Hz), ce qui peut donner l'impression que la position fréquentielle du formant serait déterminée au millième de milliardième de Hertz.

La comparaison entre logiciels montre néanmoins des différences d'une toute autre échelle entre les estimations des fréquences des formants : ces différences sont fréquemment de plusieurs dizaines de Hertz. La méthode choisie, ainsi que les paramètres fixés pour la détection (par exemple, dans PRAAT, le nombre de formants à détecter et la fréquence maximum du formant le plus élevé), ont une influence considérable sur les résultats. **La détection automatique de formants est en réalité un exercice difficile** (voir par exemple Liénard 1972, et les travaux de Stephanie Seneff sur la méthode LPC depuis les années 1970)[2]. C'est du reste pour cette raison que les systèmes de reconnaissance de la parole ne

---

[1] Vifs remerciements à Jacqueline Vaissière, Michel Scheffers, Vu-Ngoc Tuân, Christophe d'Alessandro et Shinji Maeda pour les explications fournies à ce sujet. L'exposé proposé ici n'engage bien sûr que nous.

[2] Les algorithmes de détection du type LPC *(Linear Predictive Coding)* ne parviennent pas toujours à séparer la fréquence fondamentale et le premier formant dans le cas des voyelles fermées. La variation de F_0, paramètre de *source*, est pourtant sans relation avec les paramètres du *filtre* (caractéristiques résonatoires du conduit

sont pas fondés sur la mesure des formants. Les harmoniques sont d'autant plus espacés que la fréquence fondamentale est élevée, d'où une mesure d'autant moins précise que F_0 est élevée ; pour le premier formant, en particulier, cela peut conduire à des imprécisions dans la détection, comme l'illustre la figure 1.14 (reprise de Fant 1974:38), qui représente la détection des formants pour une voyelle dont le timbre est maintenu, cependant que la fréquence fondamentale monte au cours de la voyelle, de 75 à 175 Hz, donc dans une plage de fréquence fondamentale courante en parole. Dans la région grisée, les croix (x) indiquent la position du premier formant, les cercles (o) la position de l'harmonique dominant à l'intérieur du premier formant. La variation de cette dernière peut induire en erreur les algorithmes de détection des fréquences formantiques.

Plus fondamentalement, **la généralisation selon laquelle un signal de parole correspondant à une voyelle présente un nombre constant de formants** (5 formants au-dessous de 5500 Hz environ pour une voix féminine, au-dessous de 5000 Hz pour une voix masculine) **admet des exceptions** : des formants peuvent apparaître ou disparaître par l'effet de la cavité pharyngale, de la cavité trachéale, de la cavité nasale, ou d'une cavité latérale (phénomènes rapportés par Gunnar Fant, mais non pris en compte par Kenneth Stevens [Fant 1960, Stevens 1998]). Or les algorithmes de détection automatique de formants font précisément l'hypothèse d'un nombre constant de pôles au-dessous d'un certain seuil (rappelons les valeurs qui viennent d'être indiquées, les traduisant en nombre de *pôles* plutôt que de *formants* : 10 pôles au-dessous de 5500 Hz en voix féminine, au-dessous de 5000 Hz en voix masculine). Dans certains cas, ce choix aboutit à l'étiquetage comme « formant » d'un pôle qui ne correspond pas à une fréquence formantique. Le seuil doit alors être ajusté (abaissant le nombre de pôles à détecter à 9 au-dessous de 5000 Hz, par exemple). Au bilan, nos observations suggèrent qu'**une proportion non négligeable (de l'ordre de 10%) des**

---

vocal) dans le modèle (simplifié) de Fant 1960 ; en pratique, les algorithmes LPC aboutissent parfois à l'étiquetage comme premier formant (F_1) d'un harmonique de basse fréquence et d'amplitude relativement élevée, du fait que les premiers harmoniques se trouvent proches du premier formant des voyelles fermées. Plus précisément, il existe deux méthodes de détection LPC : une méthode par autocorrélation, et une méthode par covariance. C'est dans la méthode par autocorrélation que la valeur de F_1 obtenue en sortie a tendance à suivre la valeur de F_0, du fait que la fenêtre temporelle utilisée est large (typiquement 32 ms). Plusieurs tentatives d'aménagement ont été proposées par le passé pour réduire l'influence de F_0 sur F_1 dans cette méthode, mais aucune de ces tentatives n'a abouti à des améliorations sensibles. Quant à la méthode par covariance, qui utilise une fenêtre étroite (typiquement 3.2 ms, ce qui est généralement inférieur à la période), elle est rarement utilisée à l'heure actuelle du fait d'incertitudes techniques (communication personnelle de Shinji Maeda : « 1) it requires the detection of major voice excitation points, which is difficult to do without errors, and 2) the filter identified by the covariance method can be unstable »). Il paraît donc raisonnable de recourir à la méthode par autocorrélation (qui est notamment employée par le programme FOREST, développé par Michel Scheffers, Université de Kiel : Simpson et Scheffers 1995), tout en prenant particulièrement soin de vérifier visuellement les valeurs de F_1 fournies par cette méthode. (Merci à Michel Scheffers et Shinji Maeda pour nous avoir fourni ces explications.)

valeurs données par PRAAT ne correspondent pas aux formants tels qu'ils peuvent être observés visuellement sur un spectrogramme à bande large, essentiellement du fait du nombre de pôles choisi pour la détection.

Ce constat suggérait la nécessité de se donner les moyens d'une vérification visuelle des résultats item par item, et l'utilité de comparer les résultats donnés par PRAAT à ceux d'autres logiciels.

Les diverses méthodes d'estimation des formants fonctionnent de façon profondément différente et recourent à diverses astuces techniques. Par exemple, la suppression des hautes fréquences (celles qui se situent au-delà de la région où les formants sont détectés) peut faciliter la détection des pics sur le spectre, car elle revient à lisser le signal. Le logiciel PRAAT abaisse la fréquence d'échantillonnage du signal pour la détection des formants : la fréquence d'échantillonnage est fixée au double du seuil choisi par l'utilisateur pour la fréquence formantique la plus élevée. Ainsi, si l'utilisateur fixe un seuil à 5500 Hz pour 5 formants, le signal sera rééchantillonné à 11000 Hz, de sorte que la fréquence la plus élevée qui sera conservée correspondra au seuil de 5500 Hz (d'après le théorème de Nyqvist[1]). Si sa fréquence d'origine était de 44100 Hz ou 48000 Hz (valeurs courantes en enregistrement digital), les trois quarts de la plage de fréquences se trouvent supprimés. Parmi les résonances relevées comme candidates au statut de formant, les algorithmes effectuent un choix en fonction de divers critères. Par exemple, l'algorithme FOREST supprime les valeurs dont la largeur de bande est très élevée (supérieure à ¼ de la fréquence d'échantillonnage), et effectue une vérification supplémentaire lorsque deux résonances se trouvent classées dans le même formant : si l'une d'elles a une largeur de bande élevée (supérieure à 1500 Hz), les deux résonances sont fusionnées dans le calcul : elles sont remplacées par une seule résonance ayant une fréquence calculée en moyennant les deux résonances décelées (moyenne pondérée par la largeur de bande respective des deux résonances). L'ordinateur ne commet certes pas d'erreurs, au sens où il applique strictement les algorithmes ; en revanche, ces algorithmes ne sauraient garantir une numérotation exacte des pôles détectés. Les procédures comme celle qui vient d'être exposée ont pour objectif de limiter les erreurs dans la numération des formants ; elles ne sauraient exclure tout à fait ces erreurs.

En outre, **des courbes de formants plus ou moins lisses seront obtenues selon la largeur de la fenêtre d'analyse du signal** : une fenêtre large aura pour effet de lisser la courbe des formants. Ainsi, l'algorithme FOREST donne des courbes globalement plus lisses que l'algorithme Burg implémenté par PRAAT, non à cause d'un lissage du signal (les valeurs

[1] Ce théorème, aussi appelé « théorème de rééchantillonnage Whittaker-Nyquist-Kotelnikov-Shannon », est le suivant : lors de l'échantillonnage d'un signal (par exemple lors de la conversion d'analogique à digital), la fréquence d'échantillonnage doit être supérieure au double de la largeur de bande du signal d'entrée pour que le signal original puisse être intégralement reconstruit à partir de la version échantillonnée. En formule : si B est la largeur de bande et F_s la fréquence d'échantillonnage, la condition suivante doit être remplie : $2B < F_s$.

sont calculées fenêtre par fenêtre ; aucune condition n'est mise à la continuité des valeurs consécutives), mais du fait d'une fenêtre d'analyse relativement large (plus de trois fois supérieure au pas)[1] : le recouvrement des fenêtres successives a pour conséquence une courbe d'allure lisse. D'autres algorithmes parviennent à des courbes relativement lisses en pratiquant un suivi de formants *(formant tracking)*, qui se distingue de la simple détection de formants *(formant detection)* par le fait qu'une condition de continuité entre valeurs successives est posée, qui conduit à écarter certains *bons candidats* au statut de formant au profit d'autres qui s'intègrent mieux dans la courbe générale du formant pour l'item considéré. Dans un vocabulaire familier, les auteurs d'algorithmes de détection parleront de « cuisine » ou de « bidouille » ; il peut s'agir de « bonne cuisine », de « bonne bidouille », mais **il faut garder à l'esprit que la détection n'est pas d'une précision d'un Hz ou même de 20 Hz**[2]. De plus, au plan perceptif, **la fréquence des formants n'est pas le seul paramètre qui compte dans leur perception**. Les auditeurs sont sensibles à l'amplitude relative des formants. Beddor et Hawkins plaident pour des modèles qui prennent en compte leur amplitude, ainsi que l'allure générale du spectre : « models of vowel perception that combine peak picking with some sort of whole-spectrum approach » (Beddor et Hawkins 1990 ; voir également Ito, Tsuchida *et al.* 2001). Pour des raisons pratiques, seules les fréquences formantiques ont été estimées dans la présente étude ; dans l'interprétation des résultats, il est nécessaire de garder à l'esprit qu'il s'agit de mesures relativement frustes.

2.1.3. Choix d'une procédure pour l'estimation des fréquences formantiques

La mesure des formants n'a pas donné lieu à un développement logiciel spécifique, seulement à une réflexion d'utilisateur amené à effectuer un choix raisonné d'après une comparaison des performances constatées, sans évaluer le détail de l'implémentation des divers algorithmes de détection.

2.1.3.1. Principes de fonctionnement des outils de détection des formants proposés par le logiciel PRAAT

Le logiciel PRAAT fournit une détection des formants qui repose sur une analyse LPC (Linear Predictive Coding) : pour chaque fenêtre d'analyse, PRAAT applique une fenêtre de Gauss et calcule les coefficients LPC par l'algorithme de Burg tel qu'il est donné par Childers 1978:252-255 et Press, Teukolsky *et al.* 1992 (voir le manuel de PRAAT, page « To

---

[1] Plus précisément : fenêtre de Blackman de 30 ms, ce qui correspond à une taille effective de la fenêtre de 18 ms. Le pas est de 5 ms.

[2] L'incertitude expérimentale de l'estimation des fréquences formantiques, selon les observations de Gunnar Fant, est de l'ordre de $\sqrt{[15^2 + (0.2\ F_0)^2]}$ (Fant 1974:42), soit, pour $F_0 = 100$ Hz, une incertitude de 25 Hz, pour $F_0 = 200$ Hz, une incertitude de 42 Hz. Voir également Monsen et Engebretson 1983, « The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction », et Wood 1989, « The precision of formant frequency measurement from spectrograms and by linear prediction ».

Formant (Burg) », et Boersma 2001). Une correction automatique est apportée aux valeurs ainsi obtenues : l'algorithme supprime toutes les valeurs de formants inférieures à 50 Hz, et toutes celles qui sont supérieures à la limite fixée par l'utilisateur. PRAAT offre également la possibilité d'utiliser l'algorithme de Split-Levinson (tel qu'il est implémenté par Willems : Willems 1986), mais cet algorithme est décrit dans le manuel comme « not recommended for general use ».

2.1.3.2. Choix d'effectuer une vérification visuelle des résultats

Face au constat d'erreurs de mesure, une possibilité consiste à définir préalablement la plage de variation attendue pour chaque voyelle, et à rejeter de façon automatique les mesures qui débordent cette plage. Dans le cas de travaux portant sur des corpus très étendus, l'enquêteur peut escompter que les erreurs qui demeurent après ce filtrage relativement grossier se compensent entre elles dans les moyennes, et n'empêchent pas l'émergence de tendances statistiques significatives. Par exemple, Cédric Gendrot et Martine Adda-Decker étudient les fréquences des formants des voyelles du français et de l'allemand en appliquant l'algorithme de PRAAT de façon automatisée à un grand nombre de voyelles (200 à 5000, en fonction de la voyelle considérée), sans aucune vérification manuelle des résultats : les auteurs réalisent un filtrage automatique qui exclut toutes les valeurs aberrantes par rapport à l'acoustique du conduit vocal, puis toutes les valeurs s'écartant de plus de 200 Hz des valeurs canoniques publiées par leurs devanciers pour chacune des voyelles concernées (Gendrot et Adda-Decker 2004). Au terme du filtrage, les données restantes montrent une corrélation entre brièveté de la voyelle et centralisation vocalique : plus la voyelle est courte, plus elle tend à être centralisée. Cette étude montre qu'il est possible d'obtenir des résultats cohérents par une mesure automatique suivie d'un filtrage des données, lui aussi automatisé. Grâce à la puissance de calcul des ordinateurs actuellement disponibles, le traitement de corpus de millions voire de milliards de voyelles est techniquement envisageable.

Le corpus employé dans la présente étude est de taille modeste en comparaison de ces chiffres (certains laboratoires disposent de corpus qui comptent plus de 1000 heures de parole). Le nombre total de syllabes traité pour notre expérience de comparaison entre lecture soignée et lecture insistante est de l'ordre de 5000, nombre qui impose l'automatisation d'une partie des opérations mais permet encore une vérification individuelle des mesures. **Il a donc été choisi d'effectuer une vérification visuelle des résultats, item par item**. Ce choix est relativement courant dans les études phonétiques (une telle mesure est par exemple employée dans une étude récente des voyelles du néerlandais[1] : Adank, van Hout *et al.* 2004). Un script MATLAB a été créé pour appeler automatiquement des algorithmes de

[1] L'auteur de l'article, Patti Adank, nous a signalé que les outils logiciels avaient été réalisés par le professeur Terrance M. Nearey (Université de l'Alberta) ; nous avons contacté ce dernier pour demander s'il serait possible d'obtenir ces outils, mais n'avons pas obtenu de réponse.

détection des formants et offrir une interface pour la vérification des résultats : **ces résultats apparaissent superposés à un spectrogramme à bande large du passage audio concerné, ce qui permet à l'utilisateur de se prononcer sur la précision de la détection, et d'éliminer les pics supplémentaires qui ont été étiquetés à tort comme formants** (l'interface de vérification est présentée par un exemple sur la figure 1.15 ; plus de détails sont fournis dans l'Annexe 2, section 4.2)[1]. La suppression par l'expérimentateur de valeurs jugées erronées est un compromis généralement accepté dans les études phonétiques, bien qu'elle ne soit pas sans soulever certaines questions, puisque la suppression de certaines valeurs revient à privilégier celles qui demeurent. Le détail des proportions de valeurs supprimées lors de la vérification est indiqué au paragraphe *Résultats* des expériences correspondantes ; cette proportion s'échelonne de 1.6% (locuteur anglais M6) à 17% (locuteur anglais M3), la moyenne étant de l'ordre de 6%[2]. Les figures 3.13a-b à 3.27a-b permettent une comparaison visuelle des résultats bruts fournis par PRAAT (figures 3.13a, 3.14a...) et des mêmes résultats après les suppressions (figures 3.13b, 3.14b...) ; dans la plupart des cas, ces dernières ont un effet très limité ; dans tous les cas, elles ont pour effet de limiter l'écart-type des données.

L'estimation du quatrième formant n'a pas fait l'objet d'une vérification visuelle, et les valeurs obtenues ne sont pas rapportées ici, suivant l'idée bien établie selon laquelle les formants supérieurs à F3 n'ont généralement pas de rôle linguistique distinct : « The frequencies of the three lowest formants, F1, F2, F3, are the main determinants of the phonetic quality of a vowel » (Fant 1974:5).

2.1.3.3. Emploi d'un second logiciel, pour obtenir un point de comparaison

Un second logiciel de détection de formants a été employé : FOREST (Formant Estimation), développé par Michel Scheffers (voir Simpson et Scheffers 1995)[3]. **Signalons d'emblée que**

---

[1] Une autre méthode (proposée par le logiciel SNOORI) consiste à indiquer la position approximative des formants en les dessinant à la main : l'algorithme détecte alors les pôles les plus proches des valeurs approximatives fournies en entrée par l'utilisateur. Cette méthode, qui paraît tout à fait adéquate pour notre propos, n'a pas été employée ici, pour une raison pratique : nous n'en avions pas connaissance lorsque les mesures de formants ont été réalisées.

[2] Cette proportion peut paraître relativement élevée, mais elle inclut en fait les cas dans lesquels *seule l'étiquette attachée à une valeur est modifiée, sans que cette valeur soit changée* : si un pôle supplémentaire est étiqueté comme F2 par l'algorithme de PRAAT, les valeurs qui correspondent en fait à F2 et F3 seront étiquetées comme F3 et F4 ; lorsque l'utilisateur corrige cela dans l'interface MATLAB, le simple changement d'étiquetage compte comme une modification, au même titre que l'exclusion d'une valeur indécidable. Ce mode de calcul a pour effet d'accroître artificiellement le nombre de corrections. L'examen des figures 3.13 à 3.19 permet de s'assurer visuellement de l'impact limité que les suppressions ont sur les valeurs moyennes obtenues en fin de compte.

[3] Cet outil est librement disponible à l'adresse suivante : http://www.ipds.uni-kiel.de/ms/assp/assp_DOS.zip. Le code source est disponible à : http://cvs.sourceforge.net/viewcvs.py/emu/emu/contrib/kiel

seuls les résultats fournis par PRAAT ont été retenus au final. Les principes essentiels de FOREST sont les suivants :

> « Raw resonance frequency and bandwidth values are obtained by root-solving of the Linear Prediction polynomial from the autocorrelation method and the Split-Levinson algorithm (SLA). Resonances are then classified as formants using the so-called Pisarenko frequencies (by-product of the SLA) and a formant frequency range table derived from the nominal F_1 frequency. The latter may have to be increased by about 10% for female voices. » (extrait de la documentation de FOREST)

2.2. Mesures d'intensité acoustique globale

La mesure d'intensité (amplitude globale du signal audio) ne pose pas de difficultés techniques. Nous employons le terme d'*intensité acoustique* dans le sens classique : enveloppe du signal audio (*Root-Mean-Square amplitude of the audio signal*). Pour les données d'anglais et de naxi, les valeurs sont fournies par le programme FOREST, qui outre l'estimation des fréquences formantiques réalise une mesure d'intensité. Dans le cas des données vietnamiennes, l'algorithme FOREST n'a pas été appliqué pour une raison technique (la nasalisation des voyelles dans les 42 syllabes à consonne nasale finale rend très peu fiable la détection des formants ; voir chapitre III, section 3) ; le calcul de l'intensité pour ces données est effectué dans l'environnement MATLAB, par la formule standard y = 20 * log10(abs(x)), avec une fenêtre de 20 ms. Dans tous les cas, les mesures couvrent la portion de signal audio qui correspond, pour chaque item, à l'intervalle de temps qui sépare la première fermeture glottique de la dernière (ces fermetures étant détectées sur le signal électroglottographique).

L'intensité relative a été calculée sans calibration ; la calibration, qui aboutit à une mesure en décibels absolus, est une procédure relativement contraignante et généralement considérée comme inutile pour les signaux de parole (voir par exemple Fletcher 1972).

2.2.1. Variabilité en fonction de la position du locuteur par rapport au micro

La mesure d'intensité (à la différence des mesures de formants et de F_0, par exemple) **est très sensible à la distance entre la bouche du locuteur et le micro, ainsi qu'à l'orientation de la tête du locuteur**[1]. Une solution consiste à employer un micro-casque, de sorte que la distance au micro demeure constante. Un léger inconvénient de cette méthode tient à la moindre qualité des micros-casque en comparaison des micros de plus grande taille que l'on fixe sur trépied ou suspend au plafond du studio d'enregistrement ; le son enregistré à quelques centimètres de la bouche du locuteur présente en outre un équilibre spectral différent de celui enregistré à une distance de 30 à 50 cm. Nous n'avons pas employé de

[1] Une réflexion générale sur l'intensité et sa perception se trouve chez Fletcher 1972, qui propose également des valeurs moyennes donnant un ordre d'idée de la marge de variation d'intensité en parole.

micro-casque ; la distance du locuteur au micro et son orientation ont néanmoins été contrôlés dans la mesure du possible, en indiquant aux consultants de se tenir adossés à leur chaise, et en veillant à ce que leur posture ne se modifie pas au cours de l'enregistrement[1].

La consigne de conserver une posture relativement constante a été suivie sans difficulté par les locuteurs, *modulo* certains mouvements de la tête et des mains, qui ponctuent parfois un mot important. **Les mises en garde d'usage sont néanmoins de mise pour l'interprétation des mesures d'intensité, une partie de la variation observée pouvant être attribuée à une modification de la distance du locuteur au micro**.

2.2.2. L'interprétation de la mesure d'intensité

La mesure d'intensité acoustique globale n'est pas interprétable de façon transparente, pour les raisons qui viennent d'être évoquées, et du fait que l'intensité des différents segments (et notamment des différentes voyelles) n'est pas directement comparable. Chaque phonème possède une intensité intrinsèque ; l'intensité d'un segment est en outre influencée par son contexte phonémique, par un phénomène d'intensité *co-intrinsèque* (Rossi 1971, Fletcher 1972, et références citées, dont des travaux de Kenneth Stevens).

En outre, des études perceptives déjà anciennes paraissent démontrer que l'oreille est peu sensible aux changements d'intensité, en comparaison de sa sensibilité aux variations de hauteur. Enfin, en anglais, langue la plus étudiée, allongement, augmentation d'intensité et mouvements de F_0 se concentrent dans une large mesure sur la syllabe porteuse de l'accent lexical, ce qui a pu encourager à considérer que ces trois dimensions étaient largement redondantes[2] (en français, la corrélation est moins nette : par exemple, la *continuation* se traduit par une montée de F_0 souvent accompagnée d'une décroissance d'intensité [Delattre 1966a]).

De ce fait, à l'intérieur de la triade F_0-durée-intensité, l'intensité est actuellement le parent pauvre. Pourtant, les expériences qui avaient conduit, à partir des années 1950, à prêter à l'intensité du son un rôle linguistique très secondaire contenaient en fait une erreur de méthode. Dans les expériences de manipulation de l'intensité menées par Fry 1955, 1958, Mol et Uhlenbeck 1956, Issatchenko et Schädlich 1966, l'intensité était modifiée par un changement de gain global, qui affectait également toutes les fréquences ; or dans la parole naturelle, l'augmentation de l'intensité vocale résulte de phénomènes plus complexes. Il est nécessaire de tenir compte de cette spécificité, celle de la présence ou absence d'*effort vocal*. Au plan perceptif, l'oreille sait reconnaître cet effort : un stimulus produit avec un effort vocal plus grand sera jugé plus intense, même si l'intensité globale est en fait identique dans les deux cas (Glave et Rietveld 1975, et références citées par Vaissière 2004). « L'oreille

[1] Une solution technique consisterait à maintenir la tête du locuteur en place, mais cette mesure radicale était incompatible avec les consignes qui visaient à simuler une énonciation aussi naturelle que possible.

[2] Voir les travaux de Wayne Lea en reconnaissance automatique de la parole (depuis Lea 1973).

opère une sommation des composantes spectrales différente de celle qui régit la mesure de l'intensité globale » (Rossi 1971:143). La manipulation artificielle de l'intensité acoustique globale n'est donc pas un moyen efficace de modifier un signal enregistré afin de signaler la plus ou moins grande saillance d'un mot. Il n'est pas pour autant nécessaire de renoncer à utiliser cette mesure : **la variation d'intensité acoustique est l'un des corrélats de l'effort vocal ; elle peut donc être interprétée comme indice de la force articulatoire investie par le locuteur**[1].

[1] Voir également les recherches récentes de Kochanski, Grabe *et al.* 2005 : « Loudness predicts prominence: Fundamental frequency lends little ».

Table des matières du premier volet du chapitre II (« Aperçu d'ensemble de l'intonation du naxi »)

Premier volet : Recherches sur l'*intonation syntaxique* de la langue naxi 41

1. Déclinaison et frontières : réflexion sur les notions, et observations préliminaires 42

1.1. Observations au niveau du mot 44

1.2. Observations au niveau de groupements intermédiaires entre le mot et l'énoncé 45

1.2.1. Evaluation de l'abaissement de fréquence fondamentale en fin d'unité intonative dans un récit 46

1.2.2. Deux exemples fournissant une illustration complémentaire de l'effet des frontières sur la réalisation du ton de la syllabe finale d'unité intonative 48

1.3. Observations au niveau de l'énoncé 49

1.4. Observations au niveau d'unités supérieures à l'énoncé 50

2. Réduplication et prosodie : données synchroniques et hypothèse diachronique concernant les formes réduplicatives en naxi 52

2.1. Exposition des principaux faits, et premières hypothèses sur leur origine 53

2.1.1. Bref panorama sémantique des phénomènes de réduplication en naxi 53

2.1.2. Forme phonologique de la réduplication 54

2.1.3. Aucune explication phonologique évidente ne permet de rendre compte des schémas observés 56

2.2. L'hypothèse retenue ici au sujet de l'origine du changement : phonologisation de l'effet du marquage des frontières, à l'occasion d'une transition inaboutie vers un système prosodique à ton de mot 58

2.2.1. Le constat phonétique d'un jeu sur la ligne de déclinaison 58

2.2.2. Hypothèse diachronique 59

2.3. Volet expérimental 60

2.3.1. Méthode : étude combinée de données de production et de perception 61

2.3.2. Les données de production montrent l'ampleur de la variation intonative dans la réalisation des tons (« variation allotonique ») 63

2.3.3. Les tests de perception confirment que la différence entre allotones peut, dans les conditions de l'expérience, être perçue comme une différence catégorielle 65

2.3.3.1. Première partie du test : Syllabes isolées 68

2.3.3.2. Deuxième partie du test : Paires de syllabes 69

2.3.3.3. Troisième partie du test : Quadruplets de syllabes 72

3. Approche de la ligne de déclinaison et des frontières par l'emploi d'énoncés dont toutes les syllabes portent le même ton 74

3.1. Méthode : matériau linguistique et enregistrements 75

3.1.1. Matériau linguistique 75

3.1.2. Mesures de F_0 et quotient ouvert réalisées 77

3.1.3. Représentation et analyse des résultats : résumer les courbes à deux points par syllabes n'apparaît pas adéquat, du fait de la présence de phénomènes d'allongement doublés d'un abaissement final 78

3.2. Résultats : estimation de la déclinaison, et de l'effet des frontières 80

3.3. Premières remarques en vue d'une comparaison de la déclinaison entre langues 82

Bilan général au sujet de la composante syntaxique de l'intonation en naxi 83

Chapitre II. Aperçu d'ensemble de l'intonation du naxi

L'objectif de ce chapitre est d'offrir une présentation de l'interaction entre tons et intonation en naxi, sur la base de nos propres observations et expériences. Le chapitre est divisé en un premier volet consacré à la composante syntaxique de l'intonation, et un second consacré à sa composante pragmatique.

Premier volet : Recherches sur l'*intonation syntaxique* de la langue naxi

Ce chapitre commence par une caractérisation de la notion de *déclinaison*, et un bref compte-rendu de faits naxi observés en parole continue, qui paraissent révélateurs. Les hypothèses qu'ils suggèrent sont ensuite vérifiées par deux expériences qui portent sur

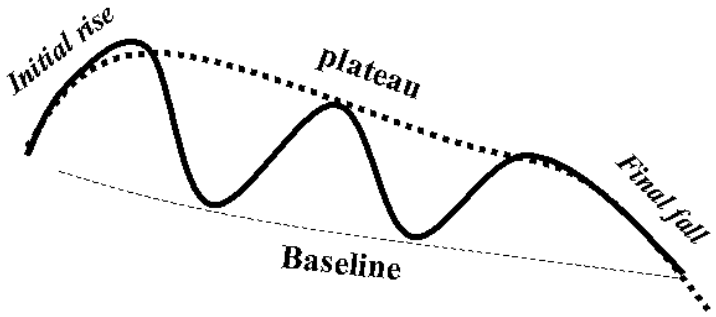
1) les schémas de réduplication (la réduplication est un procédé morphologique de création d'un composé par la répétition totale ou partielle d'un lexème ; par exemple, en naxi, /lá/, « frapper », se réduplique en /lá lɑ̄/ « se disputer »).
2) des énoncés dont toutes les syllabes portent un même ton lexical, moyen privilégié d'isoler l'effet des paramètres intonatifs.

Nos observations et expériences tendent à montrer qu'en naxi, un patron légèrement descendant s'observe (*modulo* les tons lexicaux) à l'intérieur de chacune des unités intonatives (en particulier : à l'échelle du mot, du groupe de souffle, de l'énoncé, et au-delà), s'achevant, en fin d'unité, par un allongement de la rime de la dernière syllabe et un abaissement de F_0.

La section 1 présente les notions de *déclinaison* et de *frontières*, sur un plan général, puis formule des observations sur ces phénomènes en naxi en s'appuyant sur des exemples qui paraissent éclairants, au niveau du mot (1.1), de groupements intermédiaires entre le mot et l'énoncé (1.2), de l'énoncé (1.3), et d'unités supérieures à l'énoncé (1.4). Les sections 2 et 3 visent à vérifier expérimentalement et compléter ces premières observations par l'étude expérimentale de deux questions qui nous paraissent offrir un angle d'approche privilégié : les schémas de réduplication, et des énoncés dont toutes les syllabes portent le même ton lexical.

1. Déclinaison et frontières : réflexion sur les notions, et observations préliminaires

La notion de *déclinaison* désigne la décroissance progressive de F_0 au cours de l'énoncé affirmatif, comme l'illustre la figure (très stylisée) ci-dessous, qui présente le schéma intonatif de phrase *non marqué*, ou *prototype* du schéma de phrase (Vaissière 1983).



Ligne de crête et ligne de base sont toutes deux descendantes ; la ligne de crête a une pente plus forte, de sorte que la plage de variation (écart entre ligne de base et ligne de crête) à l'intérieur de laquelle ont lieu les variations locales et semi-globales diminue à mesure de l'énoncé. La présence d'une déclinaison de F_0 au cours des énoncés affirmatifs a été constatée dans de nombreuses langues[1]. Il est établi que F_0 n'est pas un paramètre isolé ; le terme de *déclinaison* est employé par extension pour désigner la décroissance d'autres paramètres, tels que l'intensité acoustique :

> From studies of prose reading we have found a typical phrase intensity contour with an initial rise lasting about 100 ms usually followed by a declination of about 4 dB per second and a more rapid decay in the last 400 ms ending with a final abduction gesture in prepause voiced segments or with a creaky voice termination at a voiced juncture. (Fant et Kruckenberg 1995:622 ; langue étudiée : le suédois)

L'ampleur des mouvements des articulateurs connaît également une décroissance (voir par exemple Vayra et Fowler 1992, « Declination of supralaryngeal gestures in spoken Italian »).

La formulation de G. Fant et A. Kruckenberg citée ci-dessus laisse entendre que la déclinaison pourrait se quantifier de façon temporelle. W.J. Poser (Poser 1984, *The Phonetics and Phonology of Intonation in Japanese*) relève une décroissance de 10 à 12 Hz par seconde, et défend l'idée selon laquelle cette déclinaison ne serait pas commandée par des variables linguistiques. De même, J. Pierrehumbert et M. Beckman caractérisent la déclinaison comme « a backdrop phonetic process that unfolds gradually in time without

[1] L'emploi du mot remonte à Cohen et 't Hart 1967 ; le phénomène est notamment abordé par Maeda 1976 (en anglais), Cohen, Collier *et al.* 1982 (« Declination: construct or intrinsic feature of speech pitch ? ») et références citées, Ladd 1984 (« Declination: a review and some hypotheses »), Gussenhoven et Rietveld 1988.

regard to the phonological sequence of tones » (Pierrehumbert et Beckman 1988). Ce point de vue va souvent de pair avec l'idée selon laquelle la déclinaison s'expliquerait directement par des mécanismes physiologiques. Une formulation extrême apparaît chez Philip Lieberman, qui attribue la déclinaison à une diminution de pression sous-glottique au cours de l'énoncé (Lieberman 1967) ; il est établi que, toutes choses égales par ailleurs, une diminution de pression sous-glottique s'accompagne d'un abaissement de F_0. En réponse à Philip Lieberman, John Ohala souligne que la décroissance régulière de pression sous-glottique au cours de la parole ne suffit pas à expliquer la décroissance de F_0.

> Functionally the subglottal respiratory system behaves for the most part like a piston in a piston chamber driven with a constant force. (Ohala 1990:23, « Respiratory activity in speech »)

La conclusion de l'article demeure néanmoins prudente : « **It must be concluded that the question of whether F_0 declination is caused by laryngeal or by respiratory activity has still not been answered definitively.** » Helmer Strik et Louis Boves (Strik et Boves 1992, 1995) proposent un modèle dans lequel la déclinaison de F_0 peut être entièrement expliquée par la décroissance de pression sous-glottique ; ils précisent toutefois :

> ... the issue is genuinely not decidable, unless there is agreement about the way in which downtrend in fundamental frequency and subglottal pressure are defined. (Strik et Boves 1995:219)

La part que tiennent les phénomènes d'ordre physiologique dans la déclinaison n'est pas clairement établie. Divers auteurs soulignent la nécessité de données quantifiées plus abondantes sur ce sujet (Ladd 1993 ; Nolan 1995 ; Rialland 2001:301). **Il paraît néanmoins bien établi que la diminution de la hauteur mélodique à mesure de l'énoncé est intimement liée à la structuration énonciative**[1]. Les mesures de F_0 effectuées sur les langues les plus diverses établissent que la déclinaison n'est pas une constante. En thai, « l'espace tonal se rétrécit au fur et à mesure que la phrase se déroule *(et cela même dans des phrases de faible durée)* » (Gsell 1979b ; c'est nous qui soulignons) ; en yorùbá, le nombre de syllabes de l'énoncé a un effet sur la pente de F_0 des énoncés : plus l'énoncé est court, plus la pente est forte (Laniran 1992:260).

La ligne de déclinaison a part, selon des modalités variables d'une langue à l'autre, au marquage des frontières, ce qu'illustre de façon exemplaire le japonais, langue qui recourt à une remise à zéro de la ligne de déclinaison pour signaler le découpage en groupes intonatifs

[1] André Martinet le notait déjà, même si ce qu'il dit de la composante physiologique est simpliste : « … le mouvement de la courbe d'intonation est largement conditionné par la nécessité de tendre les cordes vocales en début d'émission et par la tendance économique à les détendre dès que s'annonce la fin de cette émission. Cependant, les locuteurs peuvent utiliser ce mouvement à certaines fins différentiatives selon des principes qui semblent communs à l'ensemble de l'humanité, mais sous des formes qui peuvent varier d'une communauté à une autre. » (Martinet 1960:84)

(Kubozono 1992) ; elle contribue également à séparer affirmation et interrogation ou doute (voir en particulier Thorsen 1980 ; cet ordre de phénomènes est désigné par Antoine Culioli comme le *degré de prise en charge de la relation prédicative par l'énonciateur*). Les observations de Shih Chilin et ses collaborateurs sur le mandarin, dans la perspective de la synthèse de la prosodie, vont dans le même sens (Shih Chilin 1997, « Declination in Mandarin » ; Shih Chilin 2000, « A Declination Model of Mandarin Chinese ») : dans cette langue à tons lexicaux, comme dans des langues non tonales, **le *jeu sur la ligne de déclinaison* (amplification ou annulation, remises à zéro) est porteur d'informations linguistiques**. Cela est d'autant plus évident en parole spontanée, où la ligne de déclinaison est beaucoup moins régulière que dans les énoncés *de laboratoire* (Umeda 1982, « F_0 declination is situation-dependent »).

Les diverses observations regroupées ci-dessous visent à fournir un aperçu général du marquage des frontières du mot et d'unités plus larges en naxi. La section 2 se concentre sur les **schémas de réduplication** ; la section 3 recourt à des phrases dont toutes les syllabes portent le même ton lexical (énoncés *tout au ton haut*, *tout au ton moyen*, *bas* ou *montant*) pour détailler la façon dont le marquage des frontières affecte la réalisation des tons en énoncé.

1.1. Observations au niveau du mot

Le phénomène de déclinaison est observable au niveau du mot en naxi, quelle que soit la séquence tonale. Ainsi, toutes choses égales par ailleurs, la deuxième syllabe d'un disyllabe a tendance à être réalisée à une hauteur plus basse que la première. La déclinaison paraît plus forte dans les disyllabes dont les deux syllabes portent un ton bas : par exemple, le schéma tonal du mot /çỳ nỳ/ « parfumé » (xiang[1] 香) est /LL/, tandis qu'au plan phonétique la diminution de fréquence fondamentale au cours du mot est sensible. Une telle différence phonétique (variation allotonique positionnelle) peut induire en erreur le linguiste aux premiers stades de son enquête[1].

La section 2 du présent chapitre et la section 4.2 de l'Annexe 1 exposent l'hypothèse suggérée par Martine Mazaudon (communication personnelle) selon laquelle le naxi aurait connu, au cours de son histoire, une **phase pendant laquelle il aurait entamé une évolution vers un système prosodique à *ton de mot*** (au sujet de cette notion, voir la section 3.1.5.4 du chapitre IV) ; à ce stade, l'effet de la *déclinaison* aurait joué à plein sur la réalisation

[1] Martine Mazaudon et Boyd Michailovsky nous ont très aimablement communiqué en début d'enquête leurs données personnelles sur le naxi, listes de vocabulaire collectées auprès de locuteurs originaires de la région de Lijiang. Dans leurs notations, qui sont d'orientation phonétique, comme toutes les notations d'une langue que l'enquêteur ne parle pas, le ton moyen est parfois noté moyen-descendant, essentiellement en syllabe finale de mot (monosyllabe, ou seconde syllabe de disyllabe). Ce détail phonétique, quelque peu anecdotique, rejoint les autres observations sur la réalisation des frontières intonatives **par un allongement accompagné d'un abaissement final de F_0**.

phonétique du ton de mot, qui se serait étendu sur les deux syllabes de certains disyllabes : pour prendre un exemple, la cible phonétique d'un ton L n'aurait été atteinte que vers la fin de la seconde syllabe d'un disyllabe portant ce ton. Par la suite, avant que l'évolution n'atteigne son terme (un système dans lequel un mot phonologique disyllabique ne peut porter qu'un seul ton), elle se serait inversée (peut-être sous l'influence de langues voisines). Lors du retour au système « omnisyllabique » (un ton par syllabe), chaque syllabe des disyllabes qui étaient passés au *ton de mot* se serait alors vu réassigner un ton propre ; ce processus se serait fait par une réinterprétation dans le cadre proposé par le paradigme des trois principaux tons lexicaux H, M et L ; c'est alors que **la différence phonétique de hauteur entre les deux syllabes des mots phonologiques en question aurait donné lieu à une réinterprétation** : par exemple, « pois vert, petit pois », formé de /nɣ̀/ « pois » et /hə̀/ « vert », serait un moment passé par un stade où le mot ne portait qu'un unique ton L, dont la cible était atteinte tardivement ; sa réalisation phonétique, proche de [nɣ̄ hə̀], aurait ensuite été réinterprétée comme telle au plan phonologique (d'où la forme actuelle /nɣ̄ hə̀/) lors du « retour à l'omnisyllabisme ». Si cette hypothèse se confirmait, elle montrerait la place que peut tenir la ligne de déclinaison à l'échelle du mot (et le jeu linguistique sur celle-ci) dans l'évolution de la langue.

1.2. Observations au niveau de groupements intermédiaires entre le mot et l'énoncé

Les observations au niveau de groupements intermédiaires entre le mot et l'énoncé font écho à celles effectuées au niveau du mot : une légère décroissance a lieu, s'achevant par un allongement et un abaissement plus fort.

Les observations ci-dessous proviennent d'enregistrements de parole continue (qui, dans les catégories simplifiées en usage en phonétique, seront dits de *parole spontanée*). Le nombre des variables qui changent d'un énoncé à l'autre en parole spontanée (de la composition phonémique des énoncés à l'état intérieur du locuteur) rend délicate la généralisation à partir des observations ; en particulier, les considérations de rythme gagneraient à être introduites dans la discussion. Avec ces réserves, les observations réalisées nous paraissent néanmoins livrer des indications sur le fonctionnement de la langue.

Afin que l'arbitraire de l'enquêteur ne se donne pas libre cours dans le choix des exemples à l'intérieur d'un grand corpus, il est commun de s'astreindre à relever *les n premières occurrences* d'un certain phénomène. Un premier paragraphe rapporte des mesures sur les quinze premières occurrences du possessif/relatif /gɤ̃/ dans le récit « Yulong »[1] du locuteur M3 (ce mot grammatical correspond, au plan syntaxique, à la fin d'une unité : complément du nom ou proposition relative ; du fait de l'énonciation lente et soignée de ce locuteur, une

[1] L'enregistrement intégral est joint à la thèse, dans le dossier du locuteur M3.

frontière est réalisée sur ce mot). Un second paragraphe commente deux exemples qui paraissent particulièrement évocateurs (étant bien entendu que deux exemples ne sauraient avoir valeur démonstrative).

1.2.1. Evaluation de l'abaissement de fréquence fondamentale en fin d'unité intonative dans un récit

L'impression auditive de départ, que l'expérience doit permettre de quantifier et préciser, est qu'en naxi, un abaissement de fréquence fondamentale a lieu à la fin des groupements intonatifs intermédiaires entre le mots et l'énoncé.

Les relevés rapportés ici concernent le possessif /gɤ̃/, qui est également pronom relatif, et se place après le possédant (et après la proposition qu'il transforme en relative), donc à une frontière dans l'énoncé. Le choix du possessif/pronom relatif /gɤ̃/ tient à sa fréquence élevée dans le discours ; des mesures sur cette syllabe offrent en outre l'occasion de vérifier que les mots grammaticaux connaissent en naxi une réalisation affaiblie (en comparaison de mots lexicaux de même composition phonémique et de même ton lexical), à l'instar de ce qui s'observe par exemple en français, où la fréquence fondamentale des mots grammaticaux est plus basse[1].

Selon le débit, le style de parole, et les choix énonciatifs du locuteur, une frontière intonative sera ou non réalisée. Le choix s'est ici porté sur un récit du locuteur M3 (récit auquel a été donné le titre « Yulong »). Les récits du locuteur M3 se prêtent particulièrement bien à l'étude des phénomènes d'intonation syntaxique, car son discours est très construit : énonciation professorale qui témoigne d'une longue expérience de la parole en public (en tant qu'enseignant, et dans l'administration).

Dans cette expérience-pilote, les mesures se limitent à F_0.

[1] L'observation selon laquelle le statut lexical ou grammatical d'un mot se reflète dans sa réalisation prosodique ne se vérifie pas au même degré dans toutes les langues, semblerait-il. Le phénomène serait nettement moins saillant, voire absent, dans certaines langues africaines. Cette question sera reprise dans le volet typologique de la Discussion (chapitre IV, section 3).

séquence tonale S_p S_c	position du passage concerné dans le fichier son (en sec.)	Contexte phonétique avant et observations sur la syllabe précédente (S_p = syllabe précédente)			Observations sur la syllabe-cible : le possessif/pronom relatif /gɤ̄/ (S_c = syllabe-cible)		
		syllabes précédentes, et traduction	mesures sur S_p : F_0 en début et en fin (Hz)	variation de F_0 sur S_p (%)	mesures sur S_c	variation de F_0 sur S_c (%)	rapport des valeurs finales de F_0 : (S_p / S_c) - 1, en %
H + M	265	mɤ̄ pʰɣ̣́ hə̀ kɣ̣́ gɤ̄ « où on récolte sans semer »	108-112	+4	105-72 (glottalisation finale)	(-46)	(-56)
	269	mɤ̄ lʉ̀ nà kɣ̣́ gɤ̄ « où les champs blondissent sans qu'on ait labouré »	113-116	+3	97-53 (glottalisation finale)	(-83)	(-119)
	moyenne			**+3.5**	**glottalisation finale**	**(-65)**	**(-87)**
M + M	17	nà hi¯ gɤ̄ « des Naxi »	111-106	-4.7	109-89	-22	-19
	35	ʈʂʰɯ̄ pɯ̄ gɤ̄ « de cette chanson »	113-102	-11	101-80	-26	-27.5
	57	gɤ̀ hɯ̄ gɤ̄ « d'une ascension »	86-79	-9	79-62	-27	-27
	101	kʰɯ̄ gɤ̄ « au pied de »	95-86	-10	90-40 (glottalisation finale)	-125	-115
	253	tá tɑ̄ gɤ̄ « de mariage »	115-86	-34	85-72	-18	-19
	263	ŋi¯ mā gɤ̄ « autre »	96-92	négligeable	91-69	-32	-33
	moyenne			**-11**		**-25**	**-40**
L + M	44	nʉ̀ gɤ̄ « et aussi »	83-81	négligeable	86-70	-23	-15
	102	bɣ̣̀ gɤ̄ « du dessous »	90-70	-29	75-68	-15	négligeable
	183	lē kɣ̀ gɤ̄ « et aussi »	89-75	-19	78-73	-7	négligeable
	265	mɤ̄ dò gɤ̄ « qu'on ne voit pas »	91-75	-21	80-79	négligeable	+5
	324	ŋō tá tʰʉ̀ gɤ̄	92-74	-18	74-71	négligeable	négligeable
	moyenne			**-17**		**-9**	**négligeable**
LH + M	201	ŋɑ̌ gɤ̄ « de ma famille »	79-86	+9	92-71	-30	(mesure non pertinente)
	312	nʉ̌ gɤ̄ « et aussi »	83-90	+8	100-72	-39	
	moyenne			**+8.5**		**-35**	

Tableau 2.1. Mesures sur les quinze premières occurrences du possessif /gɤ̄/ en fin de groupe dans le récit « Yulong » du locuteur M3. S_c = syllabe-cible, /gɤ̄/ ; S_p = syllabe précédente. Les cas de glottalisation finale n'ont pas été pris en compte dans les moyennes d'abaissement de F_0, pour ne pas moyenner entre mécanismes phonatoires différents.

Un net abaissement s'observe entre la syllabe qui précède /gɤ̄/ (notée S_p dans le tableau 2.4) et la syllabe /gɤ̄/ elle-même (notée S_c). Cet abaissement ne paraît pas directement lié au ton lexical de la syllabe qui précède : l'abaissement a lieu que le ton de S_p soit H, M ou LH.

Après une syllabe au ton L, le ton M de la syllabe /gɤ̄/ est réalisé à une hauteur comparable avec celle du ton L qui précède, mais s'en distingue par une fréquence fondamentale relativement stable au cours de la syllabe. Le contraste est particulièrement aisé à observer dans le cas où S_p porte, comme S_c, un ton M : la différence entre les réalisations des tons est plus grande que celle qui sépare les tons M et L enregistrés en phrase-cadre (pour lesquels des résultats détaillés sont présentés au chapitre III, section 2.2.5.1).

Les exemples au ton montant confirment l'amplitude considérable de l'abaissement final : la montée du ton haut est nettement plus faible que la descente (phonétique, non catégorielle) sur /gɤ̄/.

1.2.2. Deux exemples fournissant une illustration complémentaire de l'effet des frontières sur la réalisation du ton de la syllabe finale d'unité intonative

Les deux exemples ci-dessous fournissent une illustration complémentaire de l'effet qu'a le marquage des frontières sur la réalisation d'un ton lexical. L'ampleur de la variation allotonique positionnelle décrite ici ne surprendra sans doute en rien les spécialistes des systèmes tonals ; elle pourra en revanche être éclairante pour qui n'est pas familier de ces systèmes prosodiques, étant entendu que deux exemples isolés ne sauraient avoir valeur démonstrative.

Les deux exemples ci-dessous proviennent, comme les précédents, du récit « Yulong » du locuteur M3. Dans le passage (2), la syllabe /gɤ̄/ voisine avec une syllabe de même composition phonémique mais porteuse d'un ton lexical L : /gɤ̀/.

(2)	ŋɥ̄ lɥ̄ kɥ̄	**gɤ̀**	hɯ̄	**gɤ̄**	ʂə̄	dɯ̄	ʂə̄	wà,
	glacier	en haut	aller	POSS.	affaire, événement-	un	affaire	COPULE

« ...c'est le récit d'une ascension du glacier... »

ʈʂʰɤ̄	dɯ̄	ʂə̄	ndzə̄	nù.
ainsi	un	chose	chanter	aspect en cours

« ...voilà ce que raconte la chanson. » *(position sur l'enregistrement : à 62 secondes ; nom de l'extrait sonore correspondant : 2_A.wav)*

Le locuteur a réalisé une pause après la syllabe /gɤ̄/. **L'effet de la position est tel que la réalisation du ton M de cette syllabe /gɤ̄/, en position finale, comporte une fréquence fondamentale plus basse que celle du ton L de la syllabe /gɤ̀/** (qui, dans son emploi adverbial, signifie « en haut »). La différence entre leurs cibles finales est d'un ton musical.

Le second exemple complète le précédent, en suggérant qu'un abaissement final comparable s'observe sur les mots lexicaux et les mots grammaticaux. Dans le passage (3), la syllabe /tʰɥ̄/, en frontière de groupe, connaît un abaissement comparable à celui observé ci-dessus pour la particule possessive/relative /gɤ̄/.

(3)	ʈʂʰɯ̄	ŋi˧	kɥ́	tʰɯ̄	sé	jɤ̄,	ŋɥ̄ lɥ̄ kʰɯ̄ gɤ̄,
	DEICT.	deux	CL.	DEICT.	COND./TOP.	particule	au pied du glacier

ŋɣ̄ lɣ̣̄ bɣ̀ gɤ̃	ŋi⁻	mbē	lò	ndzʊ̀.
au-dessous du glacier	deux	village	dans	habiter

« ...et ces deux [amants], ils habitaient deux villages au pied du glacier, au-dessous du glacier. »

dʊ̄	mbē	sé	jɤ̃,	ŋi⁻ mē **tʰɣ̣̄**,	dʊ̄	mbē	sé,
un	village	COND./TOP.	particule	orient	un	village	COND./TOP.

ŋi⁻ mē gɣ̀	ndzʊ̀.
occident	se trouver

« L'un se trouvait à l'est (au levant), l'autre à l'ouest (au couchant). » *(sur l'enregistrement : à 115 secondes ; nom de l'extrait sonore correspondant : 3_A.wav)*

La rime de la syllabe /tʰɣ̣̄/ (qui signifie « sortir » ; /ŋi⁻ mē/ signifie « soleil », d'où /ŋi⁻ mē **tʰɣ̣̄**/ « orient ») est réalisée sur une fréquence fondamentale basse, et descendante.

1.3. Observations au niveau de l'énoncé

Dans nos premières notes[1], les tons sont parfois notés plus hauts que leur valeur lexicale réelle au début de l'énoncé ; l'erreur inverse apparaît en fin d'énoncé. Les erreurs sont pour l'essentiel limitées à la position initiale et la position finale d'énoncé ; de la seconde syllabe à la pénultième de l'énoncé, le cadre fourni par les syllabes environnantes semble permettre une meilleure appréciation des oppositions tonales. Les exemples (4) et (5) illustrent la perception du ton M comme ton H :

(4)	ā ki`	tʰʊ̀ !
	alcool	boire

« Bois donc un coup ! (d'alcool) » (énoncé noté au cours d'un repas, de façon erronée : /á ki⁻ tʰʊ̀/)

(5)	jū kò	kʰūī	mɤ̃	ŋgỳ !
	maison	chien	NEG.	prédicat d'existence

« Il n'y a pas de chien à la maison ! / On n'a pas de chien ! » (énoncé noté au cours d'une visite chez un voisin, de façon erronée : /já kò.../)

De façon similaire, le ton L en début d'énoncé peut être perçu comme un ton M par l'oreille du débutant, qui pratique une écoute phonétique, différente de l'écoute du locuteur natif (ou de l'apprenant avancé).

Le phénomène inverse s'observe en position finale d'énoncé : un ton H y est parfois entendu M. La particule finale interrogative /lá/ nous a d'abord paru porter un ton M : /lā/. Le fait n'est pas limité aux mots grammaticaux, ni aux énoncés longs : la troisième syllabe d'un trisyllabe dit en isolation (qui forme à lui seule un énoncé) peut être tout aussi trompeuse. Le

[1] Boyd Michailovsky (dans le cadre de la formation « Notation des langues » du LACITO) souligne le profit qu'il y a à reprendre les premières notes de terrain à une étape ultérieure : ces notes prises à un moment où la perception n'était pas encore filtrée par les cadres phonologiques de la langue contiennent des indications sur le détail des réalisations phonétiques.

mot désignant la bassine utilisée pour se laver le visage, /gi ¯ tsè pá/, a d'abord été noté /gi ¯ tsè pā/[1]. Le ton M en position finale peut également donner lieu à une identification erronée : l'unique exemple présent dans nos notes est celui d'une particule finale au ton M suivant un mot lexical au ton M, et perçue comme ayant un ton L, du fait d'un contraste intonatif (et non catégoriel) entre le ton du mot lexical et le ton de la particule ; la faiblesse relative des mots grammaticaux est bien attestée dans de nombreuses langues (voir par exemple Vaissière 1971 pour le français).

La déclinaison à l'échelle de l'énoncé sera étudiée de façon expérimentale ci-dessous (section 2.2.3), recourant à des énoncés dont toutes les syllabes portent un même ton lexical.

1.4. Observations au niveau d'unités supérieures à l'énoncé

La déclinaison à l'échelle d'unités supérieures à l'énoncé paraît comparable à celle observée dans d'autres langues, comme le français. La comparaison de F_0 au début et à la fin d'un récit de quelque longueur révèle souvent une différence aisée à observer : par exemple, le récit de l'origine des Naxi raconté par la locutrice F1 (récit « Origine », dans le dossier de la locutrice F1) commence sur une fréquence fondamentale relativement élevée ; ses dernières phrases, douze minutes plus tard, ont un débit plus lent, une intensité faible, et le passage en mécanisme phonatoire 0 y est fréquent. (Le mécanisme phonatoire 0 est communément appelé *creaky voice ;* au sujet des mécanismes phonatoires, voir Roubeau, Chevrie-Muller *et al.* 1987.)

Un aperçu du même phénomène est également fourni par les séries de syllabes enregistrées par le locuteur M1 : une syllabe lui était donnée, par exemple /tʰá/ « être possible » ; le consultant devait alors dire le mot deux fois, puis une fois en phrase-cadre (/ŋɤ̀... ʂɤ́/, « je dis... »), et ensuite répéter l'opération en faisant varier le ton, dans un ordre des tons qui était imposé :

1) dans une première série d'enregistrements : ordre H M L, soit, avec l'exemple de la syllabe /tʰɑ/ : tʰá, tʰá, ŋɤ̀ tʰá ʂɤ́ ; tʰā, tʰā, ŋɤ̀ tʰā ʂɤ́ ; tʰà, tʰà, ŋɤ̀ tʰà ʂɤ́.
2) dans une deuxième série : ordre L M H
3) dans une troisième série : ordre M L H

Ce dispositif, auquel le consultant s'est plié de bonne grâce, visait à éviter un effet systématique de la position dans la série : par exemple, une plus grande longueur des syllabes au ton L n'aurait pu être attribuée avec certitude à l'effet de ce ton si les syllabes en question se trouvaient systématiquement à la fin d'une série, position où un allongement

[1] Le fait que dans l'un et l'autre cas la voyelle soit /ɑ/, voyelle dont l'articulation a pour effet notoire d'abaisser légèrement la fréquence fondamentale, paraît être une simple coïncidence : le mot /jā kò/, qui illustre l'erreur inverse (un ton entendu plus haut que sa catégorie lexicale), comporte la même voyelle /ɑ/.

final de nature intonative est attendu. Chacune des séries forme en effet une unité, qui peut être caractérisée comme un *tour de parole* (puisque le locuteur s'interrompt à la fin de la série, et attend que l'enquêteur dise le mot chinois qui le conduira à la série suivante), ou à l'aide d'autres notions (telles que le *paragraphe oral* de Morel et Danon-Boileau 1998 ; nous n'avons pas décelé dans les données naxi de régularités qui inviteraient à postuler une hiérarchie bien définie de constituants au-delà de l'énoncé). Un effet de déclinaison à l'échelle de cette unité, très léger mais régulier, ressort en effet des mesures : un ton réalisé en fin de série est plus bas d'un cinquième de ton musical en moyenne que le même ton en début de série.

Bilan des observations générales sur la déclinaison et les frontières en naxi

L'abaissement final est aisément observable en naxi, aux divers niveaux (à l'échelle de la syllabe, du mot, du groupe de souffle, de l'énoncé, et jusqu'à l'échelle d'un récit entier). Il ne conduit pas à l'ajout de tons supplémentaires (tels que L ou M), ni à un changement catégoriel du ton de telle ou telle syllabe : par exemple, le remplacement d'un ton M par un ton L, ou un abaissement catégoriel. Le naxi ne possède pas de *downstep non automatique* (au sujet de cette notion, voir l'entrée correspondante du Glossaire). Il ne paraît donc pas utile, dans cette langue, de recourir à la notion de *tons de frontière*, sauf à préciser qu'une différence de nature sépare les tons de frontière des tons lexicaux ; mais quelle raison reste-t-il alors de choisir une appellation identique, « ton », pour les tons lexicaux comme pour les phénomènes intonatifs de frontières ? Il paraît plus avisé de parler simplement de *frontières*.

Les deux chapitres qui suivent présentent une approche expérimentale plus systématique de la question de la déclinaison et des frontières dans l'intonation du naxi.

2. Réduplication et prosodie : données synchroniques et hypothèse diachronique concernant les formes réduplicatives en naxi[1]

La réduplication est un procédé morphologique de création d'un composé par la répétition totale ou partielle d'un lexème ; par exemple, en naxi, /lá/, « frapper », se réduplique en /lá lā/ « se disputer ».

Les composés réduplicatifs du naxi, disyllabiques et quadrisyllabiques, apparaissent comme des unités particulièrement soudées, en comparaison de disyllabes lexicaux dont chaque syllabe possède un sens lorsqu'elle est considérée isolément. En ce sens, ces composés offrent un angle d'approche privilégié pour observer en synchronie l'action du découpage en constituants intonatifs. Il convient néanmoins de préciser que les conclusions ne peuvent être étendues telles quelles à la langue dans son ensemble, du fait que la réduplication appartient aux franges expressives de la langue : il serait imprudent de fonder une analyse de l'intonation *syntaxique* sur des formes qui ont autant partie liée avec l'intonation *pragmatique*. Cette section ne représente donc qu'un volet de l'étude ; elle est complétée par l'étude d'énoncés dont toutes les syllabes portent le même ton lexical (en section 3).

Les langues d'Asie du sud et du sud-est présentent une grande variété de formes réduplicatives (voir notamment Abbi 1997). Le naxi ne fait pas exception à ce constat typologique. S'agissant du classement de ces formes, dont la sémantique présente un assez large éventail, un argument phonologique précis paraît délimiter sans ambiguïté deux ensembles. En effet, un phénomène régulier de changement tonal a lieu dans deux cas (réduplication de monosyllabes portant un ton H ou un ton L) et non dans les autres (monosyllabes au ton M et disyllabes).

L'analyse instrumentale vise à montrer la proximité phonétique qui existe entre les formes avec et sans changement catégoriel de ton. Elle suggère que ces deux ensembles pourraient avoir pour origine commune une réduplication à l'identique, et que l'évolution ultérieure de deux des schémas serait due à l'effet de la ligne de déclinaison, cela dans un état de langue où le système prosodique était dans une étape de transition. L'hypothèse diachronique (suggérée par Martine Mazaudon, communication personnelle) est la suivante : lors d'un certain stade de l'histoire de la langue, le système prosodique se serait engagé vers la création de *tons de mot* en remplacement d'un système dans lequel chaque syllabe possédait son ton lexical propre ; cet épisode, sans doute relativement bref, aurait laissé pour trace les schémas de réduplication avec changement tonal (ainsi que d'autres traces, évoquées en section 4.2.1 de l'Annexe 1). L'étude de cette question sera divisée en un exposé des faits phonologiques, un exposé des hypothèses, et un volet expérimental.

[1] Une version courte de ce chapitre paraîtra prochainement dans l'ouvrage *La réduplication*, dirigé par A. Morgenstern et S. Platiel, Paris, Ophrys, coll. Bibliothèque de Faits de langues : Michaud A. et Vaissière J., « Le devenir phonétique des formes réduplicatives : réduplication et prosodie dans une langue à tons, le naxi ».

2.1. Exposition des principaux faits, et premières hypothèses sur leur origine

En naxi, de nombreux exemples de réduplication de mots lexicaux s'observent dans la langue courante (dialogues et récits), réduplication qui suit des schémas réguliers : dans certains (types A > AA et AB > AABB), un changement catégoriel de ton a lieu, dans d'autres non (type AB > ABAB). (Selon une convention répandue, dans les expressions AA, AABB et ABAB, chaque lettre désigne une syllabe : ainsi, /lá lɑ̄/ « s'affronter », issu de la réduplication du verbe /lá/ « frapper », sera décrit comme réduplication *A > AA*. Le symbole > signifie ici *se réduplique en*.) La comparaison de ces deux ensembles paraît fournir des indications importantes qui aident à cerner l'origine des schémas de réduplication actuels. Avant d'aborder cette question phonétique/phonologique, un rapide panorama des emplois des formes réduplicuées donnera une idée du rôle qu'elles jouent dans la langue naxi.

2.1.1. Bref panorama sémantique des phénomènes de réduplication en naxi

Dans cette langue, qui ne possède presque aucune morphologie verbale, la réduplication remplit un rôle de dérivation morphologique. La réduplication peut exprimer la répétition d'une action, ou la prolongation d'une activité : /fɣ̣ ná/ « rapiécer », /fɣ̣ fɣ̣ ná nɑ̄/ « s'escrimer à rapiécer [un vêtement très usé] ». Elle peut également exprimer l'activité à laquelle réfère le verbe *entendue dans un sens élargi*, incluant tenants et aboutissants : /mé/ « enseigner », /mé mē/ « assurer la formation [de qqn] » ; /tsè/ « utiliser (un outil, un objet) », /tsē tsè/ « faire plein usage de »[1] ; /ŋgwɤ̄/ « visiter, aller voir », /hjɤ̄/ « se détendre », /ŋgwɤ̄ ŋgwɤ̄ hjɤ̄ hjɤ̄/ « être oisif » (s'emploie aujourd'hui pour « faire du tourisme »). La forme réduplicuée est également attirée vers un sens figuré, par exemple dans : /pʰəɤ̄/ « dénouer [un nœud] », /pʰəɤ̄ pʰəɤ̄/ « dénouer [une situation] » :

(6)	ʈʂʰɯ̄	ʂəɤ̄	tʰɯ̄	lē	**pʰəɤ̄ pʰəɤ̄**	mɤ̄	tʰv̄.
	DEICT.	affaire	DEICT./THEM	à nouveau	dénouer (DUPL.)	NEG.	parvenir

« Cette situation était inextricable. » *(récit « Yulong » du locuteur M3, à 6 mn 47 s sur l'enregistrement ; extrait correspondant :* 6_A.wav*)*

L'emploi d'une forme réduplicuée pour les sens figurés s'observe également pour les verbes statifs : /tʰò/ « être adossé [à un arbre...] » ; /tʰō tʰò/ « être frontaliers [se dit de deux pays, par exemple] ». Lorsque leur sens s'y prête, les verbes réduplicués prennent sens de réciproque : /lá/ « frapper » > /lá lɑ̄/ « se quereller », /kʰɑ̄/ « insulter » > /kʰɑ̄ kʰɑ̄/ « s'entr'insulter », /sɯ̄/ « connaître, savoir » > /sɯ̄ sɯ̄/ « faire connaissance ; avoir une liaison ». Ce phénomène, présent dans d'autres langues, est rapporté par Alexandre François à une valeur commune à toutes les réduplications de verbes, qui signaleraient un *éclatement*

[1] Par exemple, avec pour objet /tʰē ɯ̄/, « livre, connaissance », /tʰē ɯ̄ tsè/ signifie « se servir de ses livres/ se servir de ses connaissances », et /tʰē ɯ̄ tsē tsè/ « mettre en application ses connaissances, trouver un champ d'application au savoir que l'on possède ».

du procès (François 2004:184-185). Cet éclatement peut être celui de deux ou plusieurs actants qui n'agissent pas à l'unisson, ou celui d'un procès qui ne s'oriente pas vers un accomplissement, se diluant en *activité*.

Deux exemples illustreront cette valeur de la réduplication. Dans un conte, une jeune fille qui n'arrive pas à descendre d'un arbre voit passer deux bergers, et tente d'attirer leur attention. Dans ce contexte, le verbe /ndə̀/ « appeler, crier » est réduplicqué en /ndə̄ ndə̀/ ; non pas action ponctuelle d'appeler, mais action répétée, effort soutenu : « elle se mit à crier, à les appeler ». Le second exemple (énoncé 7) est tiré d'un récit sur la façon dont on utilisait l'eau des canaux de la ville de Lijiang avant l'arrivée de l'eau courante.

(7)	ts^{h}ūū	ndzūū	gī	gȳ	hó	nè	
	repas du matin	manger	ACCOMPLI/expérientiel	passer	après	INACCOMPLI/EN COURS	
sé,			hō p^{h}é	ʈʂ^{h}ə̄ ʈʂ^{h}ə̄	t^{h}á		sè.
CONDITIONNEL/TOPICALISATEUR.			légumes	laver (DUPL.)	pouvoir, ê. autorisé		ACCOMPLI

« Quand on avait fini de manger le repas du matin, on pouvait laver ses légumes [dans le canal] » ; en d'autres termes, l'heure était venue de laver les légumes. *(enregistrement : à 1 mn 36 sec. 980 ms du récit « coutumes », dossier du locuteur M3 ; extrait sonore correspondant :* 7_A.wav*)*

En employant la forme réduplicquée, l'énonciateur décrit l'association d'une activité à un certain moment de la journée.

Enfin, certaines formes sont lexicalisées : au verbe /má mā/ « construire, réparer » ne correspond plus, dans l'état actuel de la langue, de verbe monosyllabique /má/, même si l'on peut présumer que /má mā/ provient de la réduplication d'un ancien /má/. De même, /tsē tsē/ « hacher de la viande » n'a plus de forme *simplex ;* mais s'agissant d'un geste répétitif, il est aisé de se représenter le passage d'un monosyllabe souvent réduplicqué à un disyllabe.

Dans le cas des adjectifs, les formes réduplicquées n'ont pas connu de spécialisation sémantique particulière. Le sens le plus fréquent est celui de *degré supérieur de* : ainsi, /ʈʂ^{h}ù/ « rapide » > /ʈʂ^{h}ū ʈʂ^{h}ù/ « très rapide », valeur que l'on peut décrire en termes de recentrage sur la notion, d'identification au prototype de la qualité concernée. Le choix d'une forme réduplicquée, plutôt que d'un intensif du type « très, extrêmement », véhicule une nuance stylistique *enjouée, affectueuse*.

Signalons enfin que certains noms peuvent également être réduplicqués, avec le sens de « tous les... » : /dỳ/ « sol, terre », /dȳ dỳ/ « tous lieux ; partout ».

Les exemples ci-dessus appartiennent tous au parler de la ville de Lijiang et des environs (« dialecte occidental », tel qu'il est décrit par He Jiren et Jiang Zhuyi 1985), mais les phénomènes qu'ils illustrent sont présents dans tous les dialectes naxi étudiés.

2.1.2. Forme phonologique de la réduplication

La forme phonologique de la réduplication (son *schéma tonal*, ou *gabarit tonal*) est entièrement prévisible d'après les tons de la forme *simplex*.

En naxi de AS comme dans la ville de Lijiang, le comportement du ton dans la réduplication de monosyllabes (A > AA) est le suivant : le ton moyen se réduplique à l'identique (M > MM), tandis que les tons L et H sont modifiés : L > ML, H > HM. Ce fait catégoriel, rapporté par plusieurs travaux antérieurs (He Jiren et Jiang Zhuyi 1985, He Zhiwu 1987:10-11, Guo Dalie, He Jiren *et al.* 1999:68-70), ressort clairement de la prononciation des consultants, qui savent en outre confirmer l'identité des tons (par identification avec un mot de référence : ton de /lá/ « frapper », ton de /lā/ « tigre », ton de /là/ « main », ou ton de /là́/ « bougie »)[1]. Le tableau 2.5a résume ces faits.

	ton du *simplex*	se réduplique en	
		1e syll.	2e syll.
niveaux tonals	**H**	H	
			M

	ton du *simplex*	se réduplique en	
		1e syll.	2e syll.
niveaux tonals			
	M	M	M

	ton du *simplex*	se réduplique en	
		1e syll.	2e syll.
niveaux tonals			
		M	
	L		L

Tableau 2.5a. Schèmes tonals de la réduplication de forme AA du naxi.

En naxi de FK, dans la réduplication de monosyllabes (A > AA), les tons M et H se rédupliquent à l'identique (M > MM, H > HH), tandis que les tons L et LM sont modifiés : le premier donne une séquence ML, et le second la séquence suivante : M (sur la 1e syllabe du composé réduplicqué) suivi de LM (sur la 2e syllabe).

Un grand nombre de disyllabes suivent le même schéma : dans ce cas (réduplication AABB), les tons résultants sont les mêmes que si chacune des deux syllabes était réduplicquée séparément[2].

---

[1] Une confirmation du caractère catégoriel du changement de ton est apportée par le processus d'ajout d'un ton H, qui véhicule l'*insistance* (voir la seconde partie du présent chapitre, section 2.4). Ce processus catégoriel traite les composés réduplicqués en fonction de leur ton à l'issue de la réduplication, et non du ton lexical du *simplex*. Par exemple, le prédicat /ʂʷɤ́/ « être plein, être rempli », portant un ton H, ne peut recevoir de ton H supplémentaire à des fins d'insistance ; en revanche, une fois réduplicqué en /ʂʷɤ́ ʂʷə̄/, ou, doublement, en /ʂʷɤ́ ʂʷə̄ ʂʷɤ́ ʂʷə̄/, il peut recevoir cette insistance, qui le transforme en : /ʂʷɤ́ ʂʷə̄ ˊ/, ou /ʂʷɤ́ ʂʷə̄ ˊ ʂʷɤ́ ʂʷə̄/.

[2] La seule exception à ce constat est la réduplication de disyllabes de schéma tonal MM, qui aboutit, non à MMMM, mais à LLMM. Ainsi, /bə̄ ʤʐə̄/ « en désordre » se réduplique en /bə̀ ʤʐə̀ bə̄ ʤʐə̄/ « très en désordre », /bə̄ nə̄/ « mou » en /bə̀ nə̀ bə̄ nə̄/ « tout mou », /nā hō/ « vigoureux, dru » en /nà hò nā hō/ « très vigoureux, très dru ». Certaines formes onomatopéiques sans *simplex* présentent le même schéma, par exemple /pʰà là pʰā lā/ (onomatopée qui évoque la chute de nombreux objets : « patatras »). Le schéma MM > LLMM appelle une explication différente de celle proposée ici pour les autres schémas. Peut-être le fait que *MMMM soit le seul schéma quadrisyllabique dont tous les tons sont identiques a-t-il joué un rôle : rompre l'uniformité de cette séquence revient en effet à rapprocher ce schéma des autres, par un jeu d'analogie (à un niveau relativement abstrait) ; l'exemple de la réduplication de monosyllabes M en MM montre pourtant qu'il n'existe

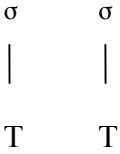
Dans un second ensemble, qui suit le schéma ABAB, un disyllabe est répliqué intégralement, y compris ses tons. Ainsi, /ndá lā/ « de petite taille » donne /ndá lā ndá lā/ « de très petite taille », /lɣ̀ fɣ́/ « tiède » donne /lɣ̀ fɣ́ lɣ̀ fɣ́/. Il peut s'agir du même simplex que dans le schéma AABB : ainsi, « très, très blanc » peut se dire /p^{h}ə̄́ sá p^{h}ə̄́ sá/ (ABAB), aussi bien que /p^{h}ə̄̄́ p^{h}ə̄́ sá sā/ (AABB). Il n'y a pas de différence sémantique qui ressorte entre l'un et l'autre schéma ; le choix paraît être une question d'habitudes personnelles. Les gabarits tonals sont identiques chez tous les consultants interrogés, tandis que les phonèmes employés, et les nuances sémantiques, varient quelque peu. Ainsi, la formule désignant un vert intense, formée sur la base de /hə̄́/, « vert-bleu », est prononcée, en fonction du consultant, /hə̄́ zá hə̄́ zá/, /hə̄́ zɯ́ hə̄́ zɯ́/, ou encore /hə̄́ kó hə̄́ kó/. Au plan sémantique, la variation est parfois importante : /pə̄́ pə̄́ hwá hwā/ signifie dans le dialecte citadin « écrire et peindre », tandis que notre principal consultant de AS (M4) emploie cette forme dans le sens « faire des graffitis (écrire sur les murs, etc.) ».

La différence phonologique entre les schémas AABB et ABAB est centrale dans la présente analyse. Des faits phonologiques exposés ci-dessus, il ressort qu'aucune modification n'a lieu si les syllabes répliquées ne sont pas en contact immédiat (schéma ABAB), tandis que dans les schémas où elles sont en contact immédiat (schémas AA et AABB), certaines séquences de tons se trouvent modifiées.

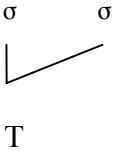
2.1.3. Aucune explication phonologique évidente ne permet de rendre compte des schémas observés

Les explications phonologiques proposées pour les schémas de réduplication de langues à tons africaines ne s'appliquent pas telles quelles aux schémas du naxi. La modification du ton d'une des syllabes dans les formes répliquées n'est pas particulière au naxi. Dans les langues africaines, plutôt que le schéma (8a), où la syllabe est répliquée avec son ton, c'est le schéma (8b) qui est couramment attesté : un même ton est associé aux deux syllabes identiques du composé répliqué. (Nous notons, très classiquement, la syllabe par σ et le ton lexical par T.)

(8a)



(8b)



Ainsi, la forme expressive bambara [bòròbárá] (qui signifie : « de façon flasque ») est analysée par Gérard Dumestre comme porteuse d'un schéma L + H, chacun des tons se déployant sur une des moitiés de la forme répliquée (Dumestre 1987:95) :

---

pas de contrainte rigide contre les séquences de tons M. Nous ne sommes donc pas en mesure de commenter plus avant le schéma LLMM.

bo ro ba ra



L H

Larry Hyman (communication personnelle) propose de décrire de la façon suivante les faits naxi, que nous lui avons exposés : le ton moyen serait un ton neutre, ton « non marqué », seuls H et L étant spécifiés lexicalement (sur la situation des tons « non marqués », situation courante dans les langues subsahariennes, voir par exemple Hyman 2001b, « Privative Tone in Bantu »). Dans la réduplication, un seul ton lexical est présent ; l'une des deux syllabes reçoit un ton M. Les tons seraient donc : H et M pour les composés issus de la réduplication d'une syllabe portant un ton H ; M et M si le ton de base était M ; L et M si le ton de base était L. Une *règle*, ou *contrainte*, imposerait ensuite un gabarit descendant sur le mot. Les principes directeurs se résumeraient de la façon suivante :

1) *Every input has to have one output.* Le ton lexical du simplex (H, M, ou L) apparaît nécessairement dans le composé réduplqué.
2) *Add a Mid tone.* (Rappelons que dans cette perspective, le ton M serait un ton non marqué.)
3) *Order the tones by descending order.*

Nous proposons ci-dessous une réflexion sur l'adéquation de cette analyse frappée au coin de l'approche autosegmentale de la tonologie africaine. Cette réflexion représente une occasion de relever des différences entre le mode d'opération de la réduplication en naxi et certains phénomènes couramment rencontrés dans le domaine africain.

Une première difficulté pour l'analyse ci-dessus concerne l'hypothèse selon laquelle le ton moyen du naxi serait un ton neutre, « sous-spécifié » : le ton M se comporte comme un ton à part entière dans le processus d'ajout de ton haut flottant (par réduction syllabique, ou par un processus intonatif d'intensification ; voir Intonation pragmatique, section 2.2.7) : une syllabe au ton M à laquelle est ajouté un ton H d'intensification vient à porter un contour montant /MH/, non un simple ton H comme ce serait le cas si la syllabe n'avait pas de ton propre au départ. Cet argument n'est cependant pas décisif en lui-même : l'idée selon laquelle M serait un ton neutre peut être défendue en distinguant plusieurs degrés dans la dérivation : il est imaginable que H et L soient seuls spécifiés lexicalement, et que les syllabes sans ton reçoivent un M *(assignation de ton M par défaut)* avant l'étape à laquelle a lieu la réassociation des tons flottants. Lorsqu'un ton H flottant s'associe à une syllabe sans ton lexical, il se poserait en réalité sur une syllabe qui a déjà reçu un ton M, d'où la création d'un contour montant MH. Cette analyse (qui s'inspire de propositions formulées par Akinbiyi Akinlabi : Akinlabi 1985, *Tonal Underspecification and Yorùbá Tone*) est contre-intuitive au vu de la diachronie du naxi, le ton haut s'étant historiquement développé sur la base d'un système qui ne comportait que deux tons, lesquels sont devenus les tons L et M du système actuel (Bradley 1975) ; pour autant, cet argument n'est pas absolument décisif au

plan synchronique. Une difficulté plus sérieuse provient du fait que le principe d'ordonnancement des tons par hauteur décroissante à l'intérieur d'un même mot *(Order the tones by descending order)* ne s'applique pas ailleurs en naxi : toute la combinatoire tonale est attestée dans les disyllabes, y compris les combinaisons /LH/, /MH/, /LM/. Toutefois, cet argument n'est pas décisif non plus, car il est imaginable que les composés réduplicatifs, appartenant aux franges expressives de la langue, aient un comportement à part : fortement intégrés (soudés), les composés réduplicatifs seraient plus enclins que d'autres unités lexicales (qui forment pareillement des *mots prosodiques*) à connaître une neutralisation tonale sur l'une de leurs syllabes, avec pour effet l'association d'un seul ton au composé entier. Une tendance à l'application d'un schéma tonal descendant à l'échelle des diverses unités intonatives s'observe bel et bien en synchronie en naxi ; il n'est pas déraisonnable d'imaginer que les schémas de réduplication de forme AA portent la trace phonologisée de cette tendance. L'hypothèse diachronique esquissée ci-dessous rejoint dans une large mesure le point de vue exprimé par Larry Hyman, quoiqu'en reformulant son idée dans un vocabulaire différent : non plus sous forme d'une règle synchronique, qui serait isolée dans l'état actuel du naxi, mais en imaginant (par analogie avec une autre langue tibéto-birmane, le tamang, décrit par Mazaudon 1973) un scénario historique dans lequel la variation allotonique (non catégorielle) a pu être phonologisée.

2.2. L'hypothèse retenue ici au sujet de l'origine du changement : phonologisation de l'effet du marquage des frontières, à l'occasion d'une transition inaboutie vers un système prosodique à ton de mot

2.2.1. Le constat phonétique d'un jeu sur la ligne de déclinaison

Un constat phonétique suggère que les schémas observés en naxi pourraient provenir de réduplications à l'identique. Ce constat phonétique est le suivant : dans les disyllabes réduplicatifs de forme ABAB, où il n'y a pas de changement de catégorie tonale, la hauteur des tons des syllabes phonologiquement identiques est très différente à l'oreille. (Dans ce qui suit, pour distinguer les réalisations phonétiques de ces quatre syllabes, elles seront désignées par A_1 B_1 A_2 B_2.) Lors de l'enquête, nos premières notations des tons oscillaient entre ${}_{L}{}^{H}{}_{L}M$ et $M^{H}{}_{L}M$, avant que les consultants ne nous reprennent, corrigeant en ${}_{L}{}^{H}{}_{L}{}^{H}$ cette interprétation.

L'hypothèse à laquelle cette observation conduit est que les schémas de réduplication AA (et AABB) du naxi proviennent d'un redoublement à l'identique ; les transformations ultérieures des séquences ^{HH} en ${}^{H}M$ (qui n'a pas encore eu lieu à FK, plus conservateur sur ce point), de ${}_{LL}$ en M_L, ainsi que de ${}_{L}M + {}_{L}M$ en $M + {}_{L}M$ en naxi de FK, ont en commun qu'elles tendent à établir un schéma mélodique descendant à l'échelle du composé réduplicatif, nivelant en tons M les tons L en début de mot et les tons H en fin de mot. Ces

transformations constitueraient la phonologisation (transformation en processus catégoriel) d'un phénomène phonétique attesté en synchronie : le marquage intonatif des frontières, lequel se manifeste à l'échelle de divers groupements, depuis le mot lexical jusqu'à l'énoncé et au-delà.

Le tableau 2.5b reprend le tableau 2.5a présenté plus haut, y ajoutant des flèches qui signalent le sens dans lequel les tons ont évolué (partant de l'hypothèse selon laquelle les schémas proviennent d'un redoublement à l'identique).

	ton du *simplex*	se réduplique en	
		1e syll.	2e syll.
niveaux tonals	**H**	H	↓
			M

	ton du *simplex*	se réduplique en	
		1e syll.	2e syll.
niveaux tonals			
	M	M	M

	ton du *simplex*	se réduplique en	
		1e syll.	2e syll.
niveaux tonals			
		M	
	L	↑	L

Tableau 2.5b. Schèmes tonals de la réduplication de forme AA du naxi, avec indication (par des flèches ↑ et ↓) des changements catégoriels qui auraient eu lieu.

Cette piste de recherche demande à ce que les conditions de la phonologisation soient précisées, faute de quoi l'hypothèse paraît gratuite : la variation allophonique est permanente dans les langues, et ne se fixe que dans des conditions particulières, déterminées par l'état du système linguistique. En effet, si une décroissance de fréquence fondamentale et d'autres paramètres a lieu en naxi à l'échelle des diverses unités intonatives, il est raisonnable d'imaginer que les locuteurs compensent perceptivement cette variation, la reconnaissance des tons prenant en compte leur position dans l'énoncé (pour anticiper, telle est en effet la tendance qui ressort de l'expérience menée ci-dessous : voir le tableau 2.11, en section 2.2.3.2).

2.2.2. Hypothèse diachronique

L'expérience menée ci-dessous vise à établir qu'il est imaginable, au plan phonétique, que les schémas de réduplication par lesquels H donne une séquence ^{H} + M, et L une séquence M + $_{L}$, proviennent, comme M + M, d'une réduplication à l'identique : ^{H} + ^{H} et $_{L}$ + $_{L}$. Le ton M s'en tire sans accrocs dans les deux dialectes ; en suivant le même raisonnement phonétique, il paraît raisonnable d'attribuer sa bonne conservation au fait qu'il se trouve au centre de la plage de hauteur : la seule condition à sa bonne reconnaissance est qu'il n'atteigne pas les valeurs *marquées* qui caractérisent les tons L et H.

Néanmoins, **la probabilité d'un tel changement ne dépend pas de la distance phonétique entre les allophones, mais de l'existence de facteurs qui, dans le système de la langue, déterminent le changement**. L'hypothèse suggérée par Martine Mazaudon (communication

personnelle) est que le naxi aurait connu, au cours de son histoire, une **phase pendant laquelle il aurait entamé une évolution vers un système prosodique à *ton de mot*** (au sujet des systèmes prosodiques dans lesquels le domaine du ton est le mot phonologique, voir la section 3.1.5.4 du chapitre IV) ; à ce stade, l'effet de la *déclinaison* aurait joué à plein sur la réalisation phonétique du ton de mot, qui se serait étendu sur les deux syllabes de certains disyllabes : pour prendre un exemple, la cible phonétique d'un ton L n'aurait été atteinte que vers la fin de la seconde syllabe d'un disyllabe portant ce ton. Par la suite, avant que l'évolution n'atteigne son terme (un système dans lequel un mot phonologique disyllabique ou polysyllabique ne peut porter qu'un seul ton, comme c'est le cas en tamang : voir Mazaudon 1973), elle se serait inversée (peut-être sous l'effet de langues influençantes). Lors du retour au système « omnisyllabique » (un ton par syllabe), chaque syllabe des disyllabes qui étaient passés au *ton de mot* se serait alors vu réassigner un ton propre ; ce processus se serait fait par une réinterprétation dans le cadre proposé par le paradigme des trois principaux tons lexicaux H, M et L ; c'est alors que **la différence phonétique de hauteur entre les deux syllabes des mots phonologiques en question aurait donné lieu à une réinterprétation**. Tel serait le cas des schémas de réduplication de type AA, mais aussi d'un certain nombre d'autres mots lexicaux : par exemple, « pois vert, petit pois », formé de /nɣ̣/ « pois » et /hə˧/ « vert », serait un moment passé par un stade où le mot ne portait qu'un unique ton L, dont la cible était atteinte tardivement ; sa réalisation phonétique, proche de [nɣ̄ hə˧], aurait ensuite été réinterprétée comme telle au plan phonologique (d'où la forme actuelle /nɣ̄ hə˧/) lors du « retour à l'omnisyllabisme ». (La section 4.2.1 de l'Annexe 1 propose une liste d'exemples.)

Dans cette perspective, il devient compréhensible que l'effet de la ligne de déclinaison ait pu se trouver phonologisé (ce que l'expérimentation phonétique, à elle seule, ne suffisait pas à établir). En effet, s'il a existé un stade de *retour vers un système prosodique dans lequel chaque syllabe possède un ton propre*, cela à partir d'un stade où certains mots ne possédaient plus qu'un ton pour le mot entier, le problème de la réinterprétation de la hauteur phonétique de chacune des syllabes d'un tel mot s'est bel et bien posé dans des termes similaires à ceux que simule l'expérience de perception mise en place ci-dessus : une réinterprétation catégorielle d'objets phonétiques au statut phonologique incertain.

2.3. Volet expérimental

L'expérience menée vise à évaluer en synchronie l'ampleur de la décroissance de fréquence fondamentale à l'échelle des disyllabes (et quadrisyllabes), tendance qui crée une tension entre catégories phonologiques et réalisations phonétiques.

2.3.1. Méthode : étude combinée de données de production et de perception

La méthode exploratoire comporte deux volets complémentaires : une quantification, au plan de la *production*, de la différence phonétique entre tons phonologiquement identiques dans le schéma ABAB, et une étude de la *perception* de cette différence.

Sept consultants (rétribués), âgés de 27 à 55 ans, locuteurs du dialecte occidental du naxi, ont enregistré les mêmes expressions réduplicuées. Il s'agit des consultants F2, M1, M4, M5, M7, M8 et M9. L'enregistrement de M1 a eu lieu en 2002 à Pékin ; l'enregistrement des autres locuteurs en 2004 dans la ville de Lijiang, où ils résident actuellement. L'enregistrement combine audio et électroglottographie ; les paramètres mesurés sont F_0, la durée, le quotient ouvert, et (pour le locuteur M1 seulement) l'intensité acoustique. (Les fichiers sonores figurent dans un sous-dossier « réduplication » du dossier des locuteurs concernés.)

Ces données de production ne pouvant être employées dans leur intégralité pour les tests de perception (pour une question de durée des tests), ce sont les données des deux premiers locuteurs qui ont été choisies ; ces données sont en outre très différentes au plan de la production (la différence entre A1 et A2, B1 et B2 étant forte chez M1, réduite chez M4). Le test de perception a consisté à présenter les syllabes isolément (40 stimuli), puis par combinaisons de deux, explorant toute la combinatoire : $A_1 + B_1$, $A_1 + A_2$, $A_1 + B_2$, $B_1 + A_1$, etc. (140 stimuli), et enfin par combinaisons de quatre (48 stimuli). Pour que l'attention des auditeurs se porte uniquement sur le ton, non sur les phonèmes ou les items lexicaux concernés, c'est le signal électroglottographique qui a été utilisé pour les tests, et non le signal audio : les auditeurs n'entendaient donc qu'une ligne mélodique portée par un bourdonnement. Les stimuli étaient présentés dans un ordre aléatoire, le même pour tous les auditeurs.

Les tests ont été soumis à 15 auditeurs, tous ayant le naxi pour langue maternelle et vivant dans la Région autonome naxi.

Certains des auditeurs ont passé le test dans la ville de Lijiang (il s'agit des locuteurs F2, M5, M7, M8, M9), les autres dans le principal lieu d'enquête : le village de /ɑ̄ ʂə̀/, abrégé en AS. Tous les auditeurs n'ont pu passer l'intégralité des tests de perception pour des raisons pratiques : les citadins participaient une ou deux demi-journées seulement à l'enquête, ce qui obligeait parfois à sacrifier une partie de notre programme de travail (qui comprenait des enregistrements et des tests de perception) ; à la campagne, les consultants étaient beaucoup plus disponibles, mais il était rare que les conditions de quiétude nécessaires au bon déroulement des tests restent longtemps réunies, l'arrivée de visiteurs ou une menue tâche à accomplir sans délai interrompant fréquemment les tests de perception. Au bilan, le nombre d'auditeurs varie de 15, pour les monosyllabes, à 10, pour les disyllabes et quadrisyllabes du locuteur M4. Les conditions de l'expérience ne permettaient pas de choisir les sujets en

fonction de critères stricts ; leur âge varie de 15 ans à 58 ans, la plupart ayant entre 30 et 55 ans.

En fonction de la tessiture de chaque locuteur, ses tons lexicaux sont réalisés à des hauteurs différentes ; afin que les auditeurs aient l'occasion de prendre la mesure de la tessiture de ce locuteur, ils entendaient au début de chacun des six tests le signal électroglottographique de rimes aux quatre tons enregistrées en phrase-cadre par le locuteur concerné.

> ... en thai, et sans doute dans d'autres langues, l'identification est meilleure si les auditeurs connaissent au préalable (par exemple grâce au contexte) la gamme, ou tessiture, de chaque locuteur : chez certains sujets, la pente des tons moyen et bas est parallèle, donc si on ne connaît pas les niveaux propres au locuteur, il est difficile d'interpréter ses tons. (Gsell 1979a:28)

La transposition dans le contexte des enquêtes de terrain des protocoles expérimentaux pratiqués en laboratoire soulève certaines difficultés spécifiques. A l'origine, les scripts MATLAB réalisés pour les tests de perception étaient conçus, comme il est courant pour les tests réalisés en laboratoire, de façon à pouvoir être passés sans que l'enquêteur soit présent : une fois le test lancé, l'auditeur suit les consignes qui lui sont fournies à l'écran. Lors du passage des tests, il était néanmoins exclu de demander aux auditeurs de passer le test seuls du fait de leur manque de familiarité avec l'informatique, qui suscitait la crainte de commettre des erreurs. De fait, l'utilisation de l'ordinateur portable représentait une responsabilité réelle : il s'agissait d'un matériel qu'ils savaient coûteux, indispensable à l'enquêteur pour son travail, et propriété d'un institut de recherche étranger. En outre, seuls quatre des consultants connaissaient la romanisation du naxi ; adopter une procédure différente pour ces auditeurs aurait nui à l'homogénéité des résultats.

C'est donc l'enquêteur qui était aux commandes du poste informatique, lisant à haute voix les consignes (affichées à l'écran en naxi, dans la romanisation décrite par He Jiren et Jiang Zhuyi 1985:131-133), et effectuant à mesure la saisie des réponses que lui communiquait verbalement l'auditeur[1]. Ce dernier pouvait se concentrer sur la tâche d'écoute. Il était assis à environ un mètre cinquante d'un haut-parleur haute fidélité. Chaque auditeur a passé le test individuellement, en l'absence de toute autre personne, l'enquêteur mis à part[2].

---

[1] Chaque saisie était systématiquement ponctuée d'un geste d'assentiment souriant, qui signalait à l'auditeur que sa réponse avait bien été entendue, et avait en outre pour fonction de convaincre l'auditeur du bien-fondé de sa réponse, et de lui donner confiance dans la sûreté de son jugement (l'utilité d'un tel comportement de la part de l'enquêteur avait été suggérée par la phase-pilote du test).

[2] Dans un unique cas, le consultant principal (M4) a insisté pour être présent lorsque son petit frère passait le test, promettant le plus grand silence. Refuser cette requête à un homme qui cumulait les titres de maître de maison, d'aîné et de consultant dévoué paraissait trop brutal, et M4 a assisté au test que passait son frère. Comme il était prévisible, il n'a pu se contenir lorsqu'il lui semblait que son frère commettait une erreur manifeste dans la reconnaissance des tons, et a multiplié les signes d'intelligence à son intention. Pour couper court à tout débat sur la *bonne réponse*, l'enquêteur a doctement annoncé que l'acoustique des lieux avait pour

La consigne du test était : imaginez que vous entendez un mot ; indiquez quel est son ton. (Pour les disyllabes et quadrisyllabes, la question était : indiquez quels sont les tons de chacune des syllabes.) La phase-pilote du test a suggéré qu'il était plus facile pour l'auditeur d'identifier le ton après avoir lui-même répété le stimulus ; il a donc été systématiquement demandé à l'auditeur de répéter ce qu'il entendait (en lui demandant d'employer la combinaison de phonèmes /pɣ/, comme l'avait fait le participant à la phase-pilote du test), puis, dans un deuxième temps, d'expliciter l'identité du ton en donnant un mot de référence, tel que, pour les monosyllabes : /lá/ « frapper », /bū/ « aller », /sò/ « apprendre », /t^hù́/ « 3^e sg. collectif »[1]. La même consigne a été donnée oralement à tous les auditeurs à partir de notes écrites.

2.3.2. Les données de production montrent l'ampleur de la variation intonative dans la réalisation des tons (« variation allotonique »)

L'examen de la figure 2.1a (locuteur M1) fait ressortir les faits suivants :

1) la fréquence fondamentale des syllabes A_1, B_1 et A_2 est significativement au-dessus des valeurs observées sur les monosyllabes en phrase-cadre, et B_2 significativement au-dessous[2].

---

effet de modifier le son, de sorte que les deux frères, assis à des endroits différents, n'entendaient ni ne pouvaient entendre les mêmes tons. Ainsi nul n'avait lieu de douter de la justesse de son audition ; l'honneur était sauf ; l'aîné, concluant à son peu d'utilité pour l'expérience, pouvait renoncer à assister sans perdre la face ; et le cadet a pu reprendre le test depuis le début.

[1] La consigne de répéter la mélodie entendue sur la syllabe /pɣ/ avant d'indiquer son identité a été répétée dans les cas où les sujets cessaient de le faire en cours d'expérience. En revanche, la méthode utilisée pour indiquer l'identité du ton n'a pas été imposée avec la même sévérité : certains consultants (F2, M4, M5, M7, M9) sont spontanément passés, au cours du test, à une désignation « métalinguistique » : haut, moyen, bas, montant, réponses que le consultant indiquait soit en naxi, soit en chinois ; le fait d'employer des mots chinois est potentiellement porteur d'un risque de *code-switching* vers cette langue, qui pourrait influencer l'écoute pratiquée ; en pratique, il semble raisonnable de penser que les désavantages de cette légère marge de liberté laissée au consultant sont minimes au regard de la contrainte de rappels multiples de la part de l'enquêteur. De plus, les consultants qui parviennent à une conscience du ton, et savent l'abstraire de la syllabe qui le porte, jugent plus satisfaisant de passer à cette étape et de fournir à l'enquêteur la réponse à sa question, plutôt que d'en rester au stade, moins valorisant, de l'identification de la mélodie du stimulus avec un mot lexical existant. Il aurait peut-être été préférable de ne faire appel qu'à des sujets ayant cette compétence ; l'examen des résultats suggère qu'il n'est pas indispensable d'épiloguer sur les particularités de ce test de perception hors laboratoire, les résultats étant assez homogènes.

[2] L'outil statistique employé est le test-t non apparié *(two-tail)*, appliqué aux courbes rééchantillonnées, qui par commodité technique comportent 100 points de données. Les valeurs de p sont inférieures à $5x10^{-2}$ sur plus de 80 des 100 points dans tous les cas.

Il importe de souligner que le choix de rééchantillonner toutes les courbes en 100 points n'implique nullement qu'une valeur en un de ces 100 points ait une quelconque valeur par elle-même, puisque dans cette division en 100 points, chaque point correspond à moins d'un cycle glottique. Le choix d'employer 100 points, plutôt que 20, 10 ou même 3 tient au souhait de ne pas dégrader la précision des mesures en amont du résultat final, de même que dans les sciences physiques les résultats numériques ne sont arrondis qu'à l'étape du résultat et non

2) A_1 est plus élevé que A_2, B_1 plus élevé que B_2.

Les figures 2.1 à 2.12 présentent, pour chacun des locuteurs, les moyennes des mesures sur les disyllabes réduplicatifs de la forme ABAB, ayant pour schéma tonal L + H + L + H : par exemple /hỳ zý hỳ zý/ « très, très rouge » ; /pʰə̀ sá pʰə̀ sá/ « très, très blanc » ; /hə̀ lə̋ hə̀ lə̋/ « très, très vert » ; /nù lú nù lú/ « très, très noir ». Ils sont formés d'un adjectif de couleur : /hỳ/ « rouge », /pʰə̀/ « blanc », /hə̋/ « vert / bleu », /nù/ « noir », suivi d'un adjectif qui oriente vers une certaine nuance de couleur et de texture. La similarité de structure de ces items autorise à les moyenner.

Sur la figure 2.1a (locuteur M1), les courbes schématiques des tons L et H tels qu'ils ont été réalisés à l'intérieur d'une phrase-cadre par le même consultant sont également reportées sur la figure (lignes noires épaisses), en vis-à-vis des syllabes concernées, offrant une valeur-repère.

Deux séries n'ont pas été enregistrées pour des raisons pratiques : celle des composés suivis d'une particule possessive /gɤ̄/ par le locuteur M1, celle des composés sans particule par le locuteur M4. Une mesure d'intensité a été réalisée pour le locuteur M1 ; elle est présentée sur la figure 2.1b.

Au plan de l'intensité acoustique globale, et de la qualité vocale telle qu'elle est reflétée par le quotient ouvert, aucune différence nette n'apparaît entre A_1 et A_2, B_1 et B_2 : la décroissance d'intensité de A_1 à A_2 et de B_1 à B_2 n'est que de l'ordre d'1 décibel, et s'explique aisément par la différence de hauteur (toutes choses égales par ailleurs, en mécanisme phonatoire 1, fréquence fondamentale et intensité ont une certaine corrélation) ; celle de quotient ouvert est également minime, et n'est pas significative.

La différence entre la hauteur de B_1 et B_2 est pareillement manifeste chez tous les locuteurs. L'allure montante de la courbe de F_0 sur la syllabe B_1 peut être interprétée comme la conséquence du caractère très élevé de la réalisation du ton H, plutôt que comme une transformation catégorielle en ton-contour montant (ce que confirment les données de perception présentées plus bas : le ton de la syllabe B_1 du locuteur M1 n'est perçu comme un ton montant que dans 7% des cas ; le même ton chez le locuteur M4 est perçu comme ton montant dans 2% des cas seulement). Les études sur l'*alignement des courbes de fréquence*

à chaque étape du calcul, ce qui introduirait en fin de calcul une incertitude élevée. Pour prendre un exemple tout théorique : soit l'opération suivante :

$[\pi * 5.3491 * 24.44842]^2 = 169349.0396$

Au terme du calcul, le résultat peut être arrondi en fonction de l'incertitude expérimentale des données fournies en entrée, et de la précision désirée. Si en revanche chaque élément est arrondi à une décimale avant l'opération, cela introduit une incertitude multipliée, au point que le résultat devient différent dès le deuxième chiffre significatif (1.7×10^5 dans le premier cas, 1.6×10^5 dans le cas où les valeurs ont été arrondies avant le calcul) :

$[3.1 * 5.3 * 24.4]^2 = 160714.3957$

fondamentale avec l'articulation, discutées dans l'Annexe 3, montrent que dans des langues variées, une cible de F_0 très élevée qui succède à une cible plus basse est atteinte tardivement, ce qui se traduit par une courbe montante sur le début de la rime syllabique concernée.

La présence d'une particule finale (/gɤ̃/, morphème de génitif) a pour effet de stabiliser la réalisation de la dernière syllabe du composé réduplicatif (B_2) : les cas de laryngalisation sont moins fréquents sur la syllabe B_2 lorsque celle-ci n'est pas en position finale avant pause. L'allure générale du composé réduplicatif est très proche dans les deux conditions d'élicitation, avec et sans particule.

La hauteur relative des tons A_1 et A_2 est plus variable d'un locuteur à l'autre. Chez M1 et M5, la rime de la syllabe A_2 a une fréquence fondamentale plus basse ; chez M7, M8 et M9, la valeur atteinte à la fin des syllabes A_1 et A_2 (cible finale) est équivalente, mais chez M9 le ton de la syllabe A_2 est plus variable que celui de A_1, et sa fréquence fondamentale moyenne est plus élevée. Le *jitter* élevé présent dans la voix de la locutrice F2 rend difficile la comparaison détaillée de A_1 et A_2. La pente plus forte de la syllabe A_2 chez tous les locuteurs peut être rapportée au fait qu'une syllabe de F_0 très élevée (B_1) intervient entre A_1 et A_2.

Au bilan, les réalisations des tons hauts sont nettement différentes selon qu'ils se trouvent en première position (B_1) ou en deuxième position (B_2) ; la différence entre les tons bas va dans le même sens (cible finale plus basse pour A_2) mais est moins forte. Ces observations laissent à penser que, dans un scénario fictif où un locuteur naxi entendrait côte à côte les tons des paires de syllabes A_1 A_2 et B_1 B_2, et ne prêterait attention qu'à ces tons (sans avoir accès au sens lexical des expressions), ces allotones d'un même ton phonologique pourraient potentiellement être perçus comme différents ; dans le cadre des catégories tonales présentes en langue naxi, la différence phonétique entre ces allotones irait dans le sens d'une perception des tons /L/ de la séquence A_1 A_2 comme M suivi de L, et de la perception des tons /H/ de la séquence B_1 B_2 comme H suivi de M. Les tests de perception qui suivent visent à vérifier cette hypothèse synchronique, qui sera ensuite mise au service d'une hypothèse diachronique.

2.3.3. Les tests de perception confirment que la différence entre allotones peut, dans les conditions de l'expérience, être perçue comme une différence catégorielle

Cette section détaille la méthode et les résultats des tests de perception ; indiquons d'emblée que ces tests confirment qu'une distance phonétique appréciable sépare les variantes positionnelles des tons dans les composés réduplicatifs, et vont dans le sens de l'hypothèse suggérée par l'examen des données de production, selon laquelle la différence entre allotones peut, dans les conditions de l'expérience, être perçue comme une différence catégorielle.

Il peut être bon de rappeler qu'il n'est pas possible de préjuger du caractère pertinent ou non, au plan de la perception, des différences constatées dans les données de production. Une étude menée par Hajek et Maeda 2000 sur la nasalité met en lumière les erreurs que risque de commettre l'enquêteur qui se fonde sur des données de production : les voyelles basses ont tendance à être produites avec un velum abaissé dans plusieurs langues (Ohala 1975, Clumeck 1976) ; pour autant, ces différences intrinsèques dans le degré d'abaissement du velum ne peuvent être considérées comme causes d'un processus de nasalisation, phonétique puis phonémique, que si elles retiennent l'attention des auditeurs ; or il apparaît précisément qu'à degré égal de couplage entre cavité oral et cavité nasale, les voyelles d'arrière sont moins fréquemment perçues comme nasales que les voyelles d'avant :

> … it has often been noted that high vowels seem to require far less velar port opening and nasal coupling to be perceived as nasal than do low vowels (...) high vowels are preferentially nasalized because, all other things being equal, low levels of acoustic perturbation favor the perception of nasalization on these vowels. (…) intrinsic differences in velic opening between vowels of different vowel heights should not be considered to be automatic causes of a differential process of nasalization (contextual and/or distinctive). (Hajek et Maeda 2000:58)

Le test de perception offre un moyen d'évaluer le degré de *compensation perceptive de la ligne de déclinaison* par les locuteurs, en leur demandant d'identifier des stimuli présentés dans diverses combinaisons.

La durée moyenne des tests est de 4, 16, et 10 mn pour les tests 1, 2 et 3 respectivement. Les tableaux 2.6a et 2.6b présentent le détail pour 12 des auditeurs ; les trois auditeurs restants n'ont participé qu'à un petit nombre de tests (auditeurs GDL, HQF, WYH).

		test 1 (une syllabe)		test 2 (deux syllabes)		test 3 (quatre syllabes)	
numéro d'audi-teur	code identifiant l'auditeur	durée en mn	nombre moyen de répétitions	durée en mn	nombre moyen de répétitions	durée en mn	nombre moyen de répétitions
1	F2	3	1.01	15	1.02	10	1.02
2	M5	non passé					
3	M7	3	1.15	17	1.10	13	1.15
4	M8	5	1.05	15	1.07	13	1.19
5	M9	3	1.02	15	1.02	10	1.02
6	M4	non passé		21 ½	1.26	10 ½	1.04
7	auditrice HQR	3	1.05	11 ½	1.04	8 ½	1.08
8	auditrice XL	3	1.03	10	1.04	14	1.19
9	auditrice HJX	2 ½	1	12	1.06	test abandonné	
10	auditeur HC	3 ½	1.02	21	1.22	non passé	
11	auditeur HSX	6 ½	1.15	26	1.40	14	1.12
12	auditeur HWJ	2	1	9	1.04	8 ½	1.23
moyennes		**3 ½**	**1.05**	**16**	**1.12**	**11**	**1.12**

Tableau 2.6a. Durée des tests et nombre de répétitions par stimulus ; stimuli du locuteur M1. La mention « test abandonné » signale un cas dans lequel l'auditrice se jugeait incapable de fournir la réponse demandée, et a souhaité interrompre la séance.

Le nombre moyen de répétitions est, pour être tout à fait précis, *le nombre moyen de fois que le stimuli a été joué*. Une valeur de 1 signifie que l'auditeur n'a jamais demandé à ce qu'aucun stimulus soit joué à nouveau, une valeur de 1.05 que l'auditeur a demandé une répétition tous les 20 stimuli en moyenne.

		test 1 (une syllabe)		test 2 (deux syllabes)		test 3 (quatre syllabes)	
numéro d'audi-teur	code identifiant l'auditeur	durée en mn	nombre moyen de répétitions	durée en mn	nombre moyen de répétitions	durée en mn	nombre moyen de répétitions
1	F2	6 ½	1.10	19	1.03	10 ½	1.02
2	M5	3	1.02	17	1.11	11	1.02
3	M7	4	1.15	12 ½	1.06	12	1.04
4	M8	5	1.05	15	1.08	10 ½	1.23
5	M9	4 ½	1.05	(35)[1]	1.05	9	1.02
6	M4	4 ½	1.05	15	1.02	11	1.19
7	auditrice HQR	4	1.05	16	1.02	7	1.08
8	auditrice XL	2 ½	1	11	1.03	6	1.08
9	auditrice HJX	5	1.03	20	1.06	non passé	
10	auditeur HC	5	1.04	12 ½	1	non passé	
11	auditeur HSX	4	1.02	35	1.22	non passé	
12	auditeur HWJ	3	1.05	15	1.09	4	1
moyennes		**4 ¼**	**1.05**	**17**	**1.06**	**9**	**1.08**

Tableau 2.6b. Durée des tests et nombre de répétitions par stimulus ; stimuli du locuteur M4.

2.3.3.1. Première partie du test : Syllabes isolées

Les syllabes extraites des séquences A_1 B_1 A_2 B_2 ont d'abord été présentées isolément. Le tableau 2.7 correspond aux données de production du premier locuteur, qui sont représentées sur la figure 2.1. Les résultats sont également repris sur le graphique 2.1 (page 491).

Le résultat majoritaire a été mis en gras dans les tableaux qui suivent, afin d'en faciliter la lecture. Ainsi, le ton bas de la syllabe A_1 du locuteur M1 a été majoritairement perçue comme un ton moyen (66% des réponses).

[1] Cette valeur est artificiellement allongée par une pause qui a dû être faite au cours du test. Elle n'est pas prise en compte dans les moyennes.

Place dans la forme réduplicuée	Ton phono-logique	Ton perçu : % de cas dans lesquels le ton a été identifié comme :				Ton produit : F_0 moyenne (en Hz), et écart (en tons musicaux) par rapport au ton M réalisé sur monosyllabe isolé
		H	M	L	LH	
1e (A_1)	L	14.5	**66**	19	0.5	106 (+ ¼ de ton)
2e (B_1)	H	**82**	10.5	0.5	7	145 (+ 1 ton ¾)
3e (A_2)	L	3	**65.5**	31	0,5	103 (même hauteur)
4e (B_2)	H	45.5	**47.5**	4	3	115 (+ 1 ton)

Tableau 2.7. Résultat du test d'identification du ton des syllabes isolées extraites des séquences réduplicuées A_1 B_1 A_2 B_2. 40 stimuli (signal électroglottographique ; données produites par le locuteur M1), 15 auditeurs.

Les deux tons L sont majoritairement perçus comme des tons M (66%) ; le premier ton H est correctement identifié (à 82%), le second crée chez les auditeurs une hésitation entre H (45.5%) et M (47.5%).

Syllabe	Ton phonologique	Ton perçu : % de cas dans lesquels le ton a été identifié comme :				Ton produit : F_0 moyenne (en Hz)
		H	M	L	LH	
1e (A_1)	L	0	32	**67.5**	0.5	129
2e (B_1)	H	**95**	3	0	2	210
3e (A_2)	L	1	29	**69**	1	128
4e (B_2)	H	**88**	**9.5**	1.25	1.25	198

Tableau 2.8. Résultat du test d'identification du ton des syllabes isolées extraites des séquences réduplicuées A_1 B_1 A_2 B_2. 40 stimuli (signal électroglottographique ; données produites par le locuteur M4), 15 auditeurs.

Les données de production du locuteur M4 (figure 2.2) montrent une articulation plus soignée que celle de M1 ; au plan de la perception (tableau 2.8 et graphique 2.2 [page 491]), cela se traduit par un taux élevé de reconnaissance du ton lexical, qui reste très majoritairement reconnu ; dans un cas sur trois, les tons L sont perçus comme des tons M, mais aucune différence nette entre A_1 et A_2 ne se dégage. Le nombre de répétitions demandées par les auditeurs fournit un indice indirect de la difficulté à identifier les tons des stimuli du locuteur M1 : une répétition demandée tous les 10 stimuli en moyenne pour les données de M1, tous les 16 stimuli pour les données de M4.

2.3.3.2. Deuxième partie du test : Paires de syllabes

Rappelons que dans la deuxième partie du test, des disyllabes artificiellement reconstitués étaient soumis aux auditeurs.

1e syllabe (ton)	2e syllabe (ton)	Ton perçu : % de cas dans lesquels le ton de la 1e syllabe a été identifié comme :				Ton perçu : % de cas dans lesquels le ton de la 2e syllabe a été identifié comme :			
		H	M	L	LH	H	M	L	LH
A_1 (L)	B_1 (H)	1.5	36.5	**62**	0	**78**	18.5	0	3.5
A_1 (L)	A_2 (L)	21	**51**	28	0	1.5	**62**	36.5	0
A_1 (L)	B_2 (H)	3	**58.5**	38.5	0	27	**61.5**	10.5	1
B_1 (H)	A_1 (L)	**88.5**	10.5	0	1	2	**83**	14	1
B_1 (H)	A_2 (L)	**86**	11.5	0.5	2	1	**73.5**	25.5	0
B_1 (H)	B_2 (H)	**87**	12	0	1	4	**88**	7	1
A_2 (L)	A_1 (L)	0	45	**55**	0	16.5	**57**	26.5	0
A_2 (L)	B_1 (H)	0.5	28	**71.5**	0	**79.5**	18.5	0	2
A_2 (L)	B_2 (H)	0	43.5	**55**	1.5	34	**61.5**	3.5	1
B_2 (H)	A_1 (L)	34	**57**	9	0	6	**75**	18.5	0.5
B_2 (H)	A_2 (L)	**53.5**	38	8.5	0	0	**70.5**	29.5	0
B_2 (H)	B_1 (H)	3	**65.5**	31.5	0	**84.25**	11.5	0	4.25

Tableau 2.9. Résultat du test d'identification du ton de disyllabes construits à partir de syllabes extraites des séquences réduplicatives A_1 B_1 A_2 B_2. 140 stimuli (signal électroglottographique ; données produites par le locuteur M1), 14 auditeurs.

Dans cette masse de résultats bruts, les lignes qui retiennent l'attention dans l'optique de la présente étude sont celles qui correspondent à :

- la séquence A_1 A_2, laquelle place côte à côte ces syllabes, pour évaluer la plausibilité phonétique d'un changement catégoriel d'une séquence L + L (dans les conditions artificielles de l'expérience, où l'auditeur ne reconnaît pas les expressions) ; et la séquence B_1 B_2, pour la même raison
- les séquences inverses, A_2 A_1 et B_2 B_1, dont la comparaison avec les précédentes offre un moyen d'évaluer l'attente perceptive que les auditeurs ont d'une décroissance de fréquence fondamentale d'une syllabe à la suivante.

Le graphique 2.3 (page 491) présente la moyenne des résultats pour les syllabes A1, A2, B1 et B2 lorsqu'elles sont en position initiale du stimulus disyllabique.

1e syllabe (ton)	2e syllabe (ton)	Ton perçu : % de cas dans lesquels le ton de la 1e syllabe a été identifié comme :				Ton perçu : % de cas dans lesquels le ton de la 2e syllabe a été identifié comme :			
		Haut	Moyen	Bas	Bas-Haut	Haut	Moyen	Bas	Bas-Haut
A_1 (L)	B_1 (H)	0	4	**96**	0	**94**	6	0	0
A_1 (L)	A_2 (L)	0	20	**80**	0	0	21	**79**	0
A_1 (L)	B_2 (H)	0	5	**95**	0	**81**	19	0	0
B_1 (H)	A_1 (L)	**96**	4	0	0	0	43	**57**	0
B_1 (H)	A_2 (L)	**93**	6	1	0	1	37	**61**	1
B_1 (H)	B_2 (H)	**89**	7	2	2	**84**	13	0	3
A_2 (L)	A_1 (L)	0	17	**83**	0	0	28	**72**	0
A_2 (L)	B_1 (H)	0	1	**99**	0	**94**	6	0	0
A_2 (L)	B_2 (H)	0	3	**97**	0	**87**	13	0	0
B_2 (H)	A_1 (L)	**90**	9	1	0	0	46	**54**	0
B_2 (H)	A_2 (L)	**87**	13	0	0	0	40	**60**	0
B_2 (H)	B_1 (H)	**78**	21	0	1	**84**	13	1	2

Tableau 2.10. Résultat du test d'identification du ton de disyllabes construits à partir de syllabes extraites des séquences réduplicatives A_1 B_1 A_2 B_2. 140 stimuli (signal électroglottographique ; données produites par le locuteur M4), 10 auditeurs.

Dans le tableau 2.10, tous les tons sont majoritairement perçus conformément à leur catégorie phonologique.

Au bilan, les principaux résultats qui ressortent du test sont les suivants :

1) concernant les séquences A_1 A_2, d'une part, et B_1 B_2, d'autre part : dans les données du locuteur M1, la séquence A_1 A_2 est perçue comme portant deux tons M (pourcentages : 51% de M pour A_1, 62% de M pour A_2), B_1 B_2 comme une séquence ^{H}M (pourcentages d'identification : 87% de H pour B_1, 88% de M pour B_2). Les stimuli extraits des données du locuteur M4 sont reconnus, conformément à leur ton lexical, comme séquences $_{LL}$ et ^{HH}, respectivement.

2) concernant les séquences inverses : la comparaison (celle de A_1 A_2 à A_2 A_1, B_1 B_2 à B_2 B_1) révèle une asymétrie. Le tableau 2.11 résume ces faits, dont le graphique 2.4 (p. 491) propose une synthèse visuelle.

position de la syllabe dans le « faux composé »	% de stimuli perçus comme ayant un ton H	% de stimuli perçus comme ayant un ton M	% de stimuli perçus comme ayant un ton L	% de stimuli perçus comme ayant un ton montant (LH)
1	31.5	38	30	0.5
2	28	56.5	14.5	1

Tableau 2.11. Pourcentages d'identifications du ton en fonction de la position dans le composé disyllabique utilisé comme stimulus. 140 stimuli (signal électroglottographique ; données produites par le locuteur M1), 14 auditeurs.

La tendance générale est celle d'un tassement des réponses concernant les tons en deuxième position de stimulus disyllabique, fréquemment entendus comme des tons M, tandis qu'en première position les oppositions sont mieux préservées : les tons L sont plus fréquemment entendus comme L, les H comme H.

position de la syllabe dans le composé	% de stimuli perçus comme ayant ton H	% de stimuli perçus comme ayant ton M	% de stimuli perçus comme ayant ton L	% de stimuli perçus comme ayant ton LH
1	44.5	9.125	46.125	0.25
2	43.75	23.75	32	0.5

Tableau 2.12. Identification du ton en fonction de la position dans le composé disyllabique utilisé comme stimulus. 140 stimuli (signal électroglottographique ; données produites par le locuteur M4), 14 auditeurs.

Une même tendance (quoique plus discrète) ressort des données du locuteur M4 (tableau 2.11, et graphique 2.5 [page 491]). La tendance attendue (une compensation perceptive de la ligne de déclinaison) se traduit principalement dans le fait que les stimuli portant un ton lexical L sont plus fréquemment entendus au ton M lorsqu'ils sont présentés en deuxième position dans l'expérience : ces données peuvent s'interpréter en disant que **l'oreille naxi s'attend à ce qu'un ton lexical M abaissé par les phénomènes de *déclinaison* et de *frontières* soit réalisé dans la plage de fréquence fondamentale d'un ton L.**

2.3.3.3. Troisième partie du test : Quadruplets de syllabes

De la troisième partie du test, dans laquelle quatre syllabes étaient présentées les unes à la suite des autres, sont seuls retenus les résultats qui concernent le quadrisyllabe d'origine, A_1 B_1 A_2 B_2. Dans les réponses concernant les données du locuteur M4, ce quadrisyllabe est perçu à 75% comme une séquence ${}_{L}{}^{H}{}_{L}{}^{H}$, conformément à ses tons lexicaux ; les 25% d'autres réponses sont ${}_{L}M_{L}M$ (et, dans un seul cas, ${}_{L}M_{L}{}^{H}$, réponse surprenante que nous attribuons à la fatigue ou l'inattention de l'auditeur). Dans les réponses concernant les données du locuteur M1, ce quadrisyllabe est perçu comme une séquence $M^{H}{}_{L}M$ dans 43% des cas ; comme ${}_{L}{}^{H}{}_{L}M$ dans 43% des cas ; comme ${}_{L}{}^{H}MM$ dans 14% des cas. Autrement

dit, **dans 100% des cas les deux tons H sont perçus, le premier comme H, le second comme M** ; la situation du ton L est plus contrastée : **les deux tons bas sont entendus à proportion égale comme une séquence LL** (dans 43% des cas**) ou une séquence ML** (dans 43% des cas également) ; enfin, dans 14% des cas ils sont perçus, le premier comme L, le second comme M. Rappelons qu'à l'audition du signal audio, qui permet de reconnaître l'expression (d'accéder à son sens), les auditeurs attribuent unanimement à ces expressions une séquence tonale $_{L}H_{L}H$.

Les résultats essentiels de l'expérience sont les suivants :

Conformément à l'hypothèse suggérée par l'examen des données de production, la différence phonétique entre les deux tons H (ceux des syllabes B_1 et B_2) est perçue par les auditeurs comme une différence de catégorie tonale. Celle-ci correspond précisément au schéma phonologique observé dans la réduplication AA du naxi occidental (H donnant une séquence H + M).

En revanche, les tons L (syllabes A_1 A_2) présentés côte à côte sont perçus comme une séquence M + M, non M + $_{L}$. Notre interprétation est que les auditeurs retiennent la plage de F_0 relativement élevée dans laquelle ces tons L sont tous deux réalisés, non la différence de hauteur entre les deux syllabes : les données de production montrent que ces tons sont réalisés plus haut que les tons L en phrase-cadre des mêmes locuteurs. Ceci amène à reconnaître que **le protocole expérimental choisi ne fournit pas une simulation exacte des conditions d'évolution d'un composé réduplicatif L > L + L** : dans l'expérience, le premier ton L est en position initiale d'un quadrisyllabe, et le rehaussement qui en résulte est très vraisemblablement supérieur à ce qui s'observerait en position initiale d'un disyllabe ; le second ton L n'est pas en position finale, et subit donc lui aussi un certain degré de rehaussement. **Le résultat confirme néanmoins l'intuition de départ selon laquelle la différence phonétique entre allotones peut, dans les conditions de l'expérience, être perçue comme une différence catégorielle.**

Le nombre restreint d'exemples en naxi ne paraît pas invalider, en lui-même, l'hypothèse diachronique proposée ici[1]. Celle-ci n'est pas invraisemblable, même s'il convient bien sûr

[1] Les exemples de disyllabes naxi dont un ton est modifié (composés réduplicatifs de schéma AA, et mots lexicaux décrits en section 4.2.1 de l'Annexe 1) rappellent, dans le domaine morphosyntaxique, les composés verbaux du type de *maintenir* en français, type « peu abondant, improductif, fait en partie de survivances, quoique certains de ses représentants soient fort usités » (Benveniste 1966 [1974]:104-105). Les exemples sont les suivants : *bouleverser, chavirer, chantourner, colporter, culbuter, maintenir, manœuvrer, morfondre, saupoudrer*, et, à l'état de dérivés nominaux, *vermoulu, saugrenu, saupiquet*. Un nom entre en composition avec un verbe qu'il précède à titre de déterminant instrumental. Benveniste relève les profondes similarités du procédé avec des phénomènes observés dans des langues amérindiennes, et conclut : « Cette création romane que nous considérons en français, quoiqu'elle se soit épuisée assez vite, demeure comme le témoignage d'une innovation typologique de grande portée générale » (*ibid.*, p. 112).

de rester prudent, notamment dans les conclusions tirées de l'expérience de perception décrite plus haut : en effet, il est quelque peu hasardeux d'imaginer simuler, par une recomposition de monosyllabes dont chacun possède un ton propre, les conditions dans lesquelles un ton de mot qui se serait étendu sur deux syllabes aurait pu être réinterprété, dans un état antérieur de la langue.

3. Approche de la ligne de déclinaison et des frontières par l'emploi d'énoncés dont toutes les syllabes portent le même ton

Il paraissait souhaitable de vérifier les observations effectuées sur les schémas de réduplication naxi par la prise en compte d'un ensemble plus vaste de phénomènes : les formes réduplicuées appartiennent aux franges expressives de la langue, elles ne sont pas nécessairement représentatives de la langue dans son entier. **Il paraît en outre souhaitable d'obtenir des données qui offrent le moyen de comparer l'influence de la déclinaison et des frontières sur les divers tons lexicaux du naxi : tons H, M, L, et LH**. Le biais employé ici pour ce faire est de recourir à des énoncés dont toutes les syllabes portent le même ton.

Dans les études de tonologie africaine, il est commun de construire des phrases répondant à des schémas tonals fixes : par exemple, des phrases dont toutes les syllabes portent un ton lexical identique, ou dans lesquels les tons H alternent avec des tons L (voir par exemple Hombert 1977, Laniran 1992, Laniran et Clements 2003). Les études de ce type sont moins courantes dans le domaine des langues d'Asie[1], dans lesquelles l'attention se concentre plus souvent sur la coarticulation tonale entre tons adjacents (Han et Kim 1974, « Phonetic variation of Vietnamese tones in disyllabic utterances » ; Xu Yi 1997, « Contextual tonal variations in Mandarin ») et sur l'effet du statut informationnel d'une syllabe sur la réalisation de son ton (Jin Shunde 1996 ; Xu Yi 1999, « Effects of tone and focus on the formation and alignment of F_0 contours »). Lors d'une expérience dont le propos se limitait à l'étude de la réalisation phonétique de diverses séquences tonales en mandarin, Xu Yi remarque qu'il n'est pas praticable de faire abstraction de la structure informative de l'énoncé :

> ... subjects often voluntarily put a focus into a sentence they were reading aloud, sometimes to correct a mistake, sometimes to make a contrast to the previous sentence, and sometimes for no apparent reason. It is therefore possible that speakers always use some kind of focus pattern when saying a sentence. If they are not told explicitly which pattern to use (...), speakers may simply pick one themselves. It thus seems advisable that focus be directly controlled (...). (Xu Yi 1997:58)

[1] Signalons néanmoins les travaux de Shen 1990 et Nguyen Thi Thanh Hoa et Boulakia 1999, qui recourent à des énoncés *tout au même ton*, choisissant le ton le moins marqué du paradigme (en vietnamien, le ton A1), afin de mieux faire ressortir les phénomènes intonatifs.

La relative rareté des expériences de manipulation systématique des séquences tonales dans le domaine asiatique nous paraît refléter, de façon indirecte, une différence dans la structuration tonale de ces deux ensembles de langues à tons, les langues d'Afrique se caractérisant par des phénomènes de *calculs de hauteurs relatives en fonction de la séquence des tons*, les langues d'Asie par la malléabilité des tons, dont la forte variation en énoncé (variation de hauteur, de longueur et d'autres paramètres) n'est pas entièrement conditionnée par leur environnement tonal. **Il est d'autant plus souhaitable de tenter l'application au naxi de procédures employées pour des langues d'Afrique, afin de préciser cette différence à la lumière des résultats publiés sur d'autres langues à tons ponctuels**, et fournir ainsi matière à une réflexion typologique (qui sera développée dans la Discussion, section 3.3). **L'enjeu de l'expérience est de faire ressortir la présence en naxi de modifications phonétiques locales non déductibles de la séquence des tons lexicaux.**

3.1. Méthode : matériau linguistique et enregistrements

3.1.1. Matériau linguistique

L'ouvrage sur le naxi de He Jiren et Jiang Zhuyi se trouve contenir deux phrases dont toutes les syllabes sont au ton M (He Jiren et Jiang Zhuyi 1985, pp. 47 et 53). Pour les tons L, H et LH, des phrases ont été préparées par l'enquêteur puis discutées avec le consultant principal (M4)[1] du dialecte de AS, et à nouveau avec les quatre autres locuteurs qui ont enregistré ces

---

[1] Le consultant principal, M4, s'est pris au jeu des combinaisons tonales : suite à notre travail sur les énoncés tout au même ton, il a spontanément créé des phrases au schéma tonal H + M + L + R, R + L + M + H, H + H + M + M + L + L + R + R, et R + R + L + L + M + M + H + H :

pɔ̠ pɔ̄	sò	là́ !	
écrire (/pɔ̠/, redupliqué en /pɔ̠ pɔ̄/)	étudier	particule signalant l'*invitation à faire qqch*	

« Tu devrais apprendre un peu à écrire ! / Tu devrais te mettre à l'écriture ! »

wɑ̀́ wɑ̀	bē	ʂɤ́
vrai (/wɑ̀/, redupliqué et avec ajout d'un ton H d'insistance)	particule adverbialisante	parler

« parler vrai / dire la vérité / dire qqch de très juste »

mɯ́	tʷɔ̠́,	ŋɖʐ̪̄	mɤ̄	pʰjɤ̀	mɤ̀,	hò́	mɑ̀́ !
vannerie	tresser	copain NEG.	aimer	particule affirmative		correct, bon	particule

« Quand on fait de la vannerie, on est mieux sans les copains, n'est-ce pas ! » (littéralement « on n'aime pas les amis ») Le consultant glose de la façon suivante : quand on fait de la vannerie, il vaut mieux ne pas être au coude à coude, serrés les uns contre les autres, sinon on ne peut pas travailler à son aise.

pî́	mɤ̀́	hà	tsʰɯ̀,	tʰē ū	pɔ̠	tʰá !	
stylo (emprunt : bi[3] 笔)	encre (emprunt : mo[4] 墨)	acheter	arriver	livre	écrire	être possible	

« Achète-toi donc un stylo et de l'encre, et tu pourras écrire ! »

Il est plus courant de dire /hà́ tsʰɯ̀/, le ton H étant dû à une syllabe /pú/ en ellipse ; à l'enregistrement, le locuteur n'a pas manqué d'ajouter ce ton : même après qu'il a appris à identifier à coup sûr les quatre tons lexicaux, il n'est pas conscient des processus par lesquels un ton H s'ajoute à une syllabe portant un ton lexical L ou M. Ces phrases ne sont pas analysées ici.

phrases (F2, M5, M7 et M9) ; un contexte d'énonciation a été imaginé, gardant à l'esprit la nécessité de parvenir à des énoncés dont l'intonation soit aussi neutre que possible, sans émotion ni effets pragmatiques saillants. Lorsque le besoin de symétrie expérimentale entrait en conflit avec les propositions du consultant, le choix a été de retenir la forme qui avait la faveur du consultant. Ainsi s'expliquent les quelques asymétries des données enregistrées : l'énoncé au ton H a huit syllabes, les autres sept ; l'énoncé au ton LH comporte deux variantes, dont l'une a été enregistrée par M4 et M5, l'autre par F2, M7 et M9. (Les enregistrements figurent dans un sous-dossier « énoncés à ton constant » du dossier des locuteurs concernés.)

Dans les notations ci-dessous, les symboles ǀ et ‖ ont été employés pour signaler les emplacements où une frontière peut être réalisée par le locuteur.

Ton haut

(9)	kó só	sé ‖	ŋ̍	kó ǀ	
	après-demain	COND./TOP.	2e sg. honorifique	particule marquant l'objet	
ʂɤ́	jɤ́	tʰɑ́.			
dire	donner	pouvoir			

« Après-demain, [je] pourrai vous le dire. » 后天就可以告诉您。

L'énoncé serait plus poli avec une particule finale d'accompli /sè/ (proche du chinois le^{0} 了 : 后天可以告诉您了), et serait alors adapté à un contexte dans lequel une des personnes s'impatiente, et l'autre l'invite à la patience. Sans la particule, la phrase revient à définir précisément le moment où l'information pourra être communiquée : *je pourrai vous le dire après-demain et pas avant*.

L'énoncé dit par M7 comporte deux syllabes portant d'autres tons : le conditionnel/topicalisateur est prononcé /sè́/ dans le dialecte de ce locuteur (NL) ; le pronom honorifique de deuxième personne /ŋ̍/ est remplacé par la simple deuxième personne /nù/ (syllabe qui devient /nū/ avant la particule /kó/). De ce fait, cet énoncé n'a pas été inclus dans les moyennes.

Ton moyen

(10)	tʰū ‖	ʈʂʰū ŋi⁻ ‖	kʰū	ʂū	mɤ̄	ndzū.
	3e sg.	aujourd'hui	chien	viande	NEG.	manger

« Il ne mange pas de viande de chien aujourd'hui. » 他今天不吃狗肉。

Par tradition, les Naxi ne mangent pas de viande de chien ; le contexte imaginé est celui d'un repas au restaurant : deux clients en observent un troisième qu'ils ont souvent vu manger de la viande de chien ; l'un remarque : « Aujourd'hui, il ne mange pas de viande de chien. » (Cette phrase est adaptée d'un exemple de He Jiren et Jiang Zhuyi 1985:47.)

Ton bas

(11) ŋɤ̀ | kʰɣ̀ kɣ̀ kò | tʰè bɣ̀ mɤ̀.
1e sg. dedans ASP. en cours ramper particule affirmative

« Je suis en train de me glisser à l'intérieur. » 我正在往里面爬。

Ton montant. La rareté des mots au quatrième ton ne permet pas de constituer un énoncé *tout au ton montant* dans lequel tous les contours montants soient lexicaux. Les syllabes /ŋɑ̋/, /nɑ̋/, /tʰɑ̋/, /hɔ̋/ ont un ton montant lexical. En revanche, le ton lexical des autres syllabes est bas : /lɣ̀/ devient [lɣ̋] par l'effet d'un /˝ sé/ en ellipse, qui laisse pour seule trace son ton H ; celui-ci crée un contour montant sur la syllabe précédente. La conjonction /nɯ̀/ « et » et l'intensif /gjà/ « très » portent fréquemment un ton H d'intensification ; le processus n'est pas gouverné par le contexte phonologique ni morphosyntaxique, et constitue un choix stylistique de l'énonciateur, de sorte qu'**il est malaisé d'imposer la réalisation d'un ton montant sur chacune des occurrences**. La solution retenue a consisté à écarter les énoncés qui ne comportaient pas de ton montant sur chacune des syllabes.

L'énoncé (12), élaboré avec M4, également enregistré par M5, n'a pas satisfait F2, M7 et M9, pour qui cet énoncé a été modifié en (12').

(12) tʰɑ̋ lɣ̋ ‖ nɑ̋ nɯ̋ ŋɑ̋ | gjɑ̋ hɔ̋ !
3e sg. COLL.[1] voir, considérer 2e sg. COLL. et 1e sg. COLL. très correct, adéquat

« Il pense que toi et moi, ça fait tout à fait l'affaire ! / il pense que notre association, à toi et moi, c'est tout à fait bien ! » 他觉得我们两个很好啊！

(12') tʰɑ ‖ nɑ̋ nɯ̋ ŋɑ̋ | gjɑ̋ hɔ̋ mɑ̋ !
3e sg. COLL. 2e sg. COLL. et 1e sg. COLL. très correct, adéquat particule

« Lui, toi et moi, ça fait tout à fait l'affaire ! / Notre association, lui, toi et moi, c'est tout à fait bien ! » 他，你和我真合适啊！

Chaque phrase a été élicitée par le biais de sa traduction chinoise. Deux répétitions par phrase étaient prévues pour chaque locuteur. Le consultant M4, plus disponible, a réalisé deux fois l'enregistrement. Le nombre de syllabes moyennées dans les figures 2.13 à 2.28 varie donc d'un locuteur à l'autre, comme l'indique le tableau 2.13.

Consultant	F2	M4	M5	M7	M9	**Total**
Nombre de syllabes	72	123	80	58	51	**384**

Tableau 2.13. Nombre de syllabes moyennées dans les calculs sur les énoncés à ton constant.

3.1.2. Mesures de F_0 et quotient ouvert réalisées

Les mesures de F_0 et quotient ouvert à partir du signal électroglottographique ont été réalisées sur l'intégralité de la rime des syllabes concernées. La mesure d'intensité

[1] Les formes désignées ici comme *COLL.*, COLLectives, font référence à la personne considérée dans son contexte familial : *moi et ma famille (mon clan)*, ou *moi et les miens*.

acoustique globale n'a pas été effectuée, la comparaison n'étant pas praticable entre voyelles dont le timbre et la longueur diffèrent.

Comme il a été indiqué au chapitre I, section 1.2.3, le quotient ouvert ne peut être calculé lorsque les pics d'ouverture ne ressortent pas clairement sur la dérivée du signal électroglottographique. La proportion de cycles glottiques dont le quotient ouvert n'a pu être calculé est indiquée, pour chaque ton et chaque locuteur, dans le tableau 2.14.

Locuteur	Ton H	Ton M	Ton L	Ton LH	moyenne
F2	13	24	19	68	31
M4	0.1	4.9	5.7	1.8	3.1
M5	0.5	5.9	0.6	2.2	2.3
M7	0	1.1	0	0	0.3
M9	0.4	2.3	0.5	0	0.8
moyenne	**2.8**	**7.6**	**5.2**	**14.4**	**7.5**

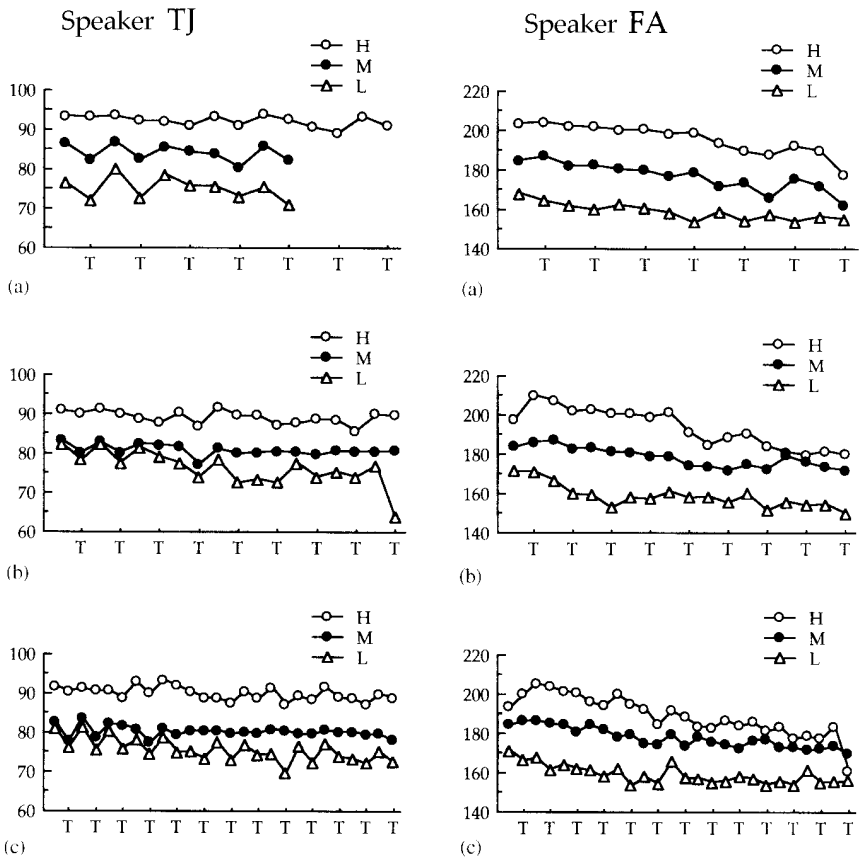
Tableau 2.14. Proportion de cycles glottiques dont le quotient ouvert n'a pu être calculé dans les énoncés à ton constant *(en %)*

3.1.3. Représentation et analyse des résultats : résumer les courbes à deux points par syllabes n'apparaît pas adéquat, du fait de la présence de phénomènes d'allongement doublés d'un abaissement final

Dans son étude de la réalisation phonétique des tons du yorùbá, Yetunde Laniran relève F_0 en deux points par syllabe :

> In this study, two measurement points are selected in each syllable, a point towards the beginning of the vowel portion of the syllable and another towards the end. (...) If instead of choosing two measurement points in each vowel, only one point were selected, we would not be able to model the appropriate F_0 pattern for each tone because (...) most level tones are realized with a falling or rising F_0 pattern (...). (Laniran 1992:60)

Ces données sont représentées sur des graphiques où les syllabes sont également espacées entre elles, ce qui revient à les représenter selon une durée normalisée, comme le montre les figures ci-dessous (reprises de Laniran et Clements 2003:220). Les figures du haut correspondent à des énoncés de sept syllabes au maximum, celles du milieu à des énoncés de neuf syllabes, celles du bas à des énoncés de treize syllabes.



From Laniran and Clements 2003:220.

Superimposed graphs of all-H, all-M, and all-L sentences for TJ and FA, showing differing rates of declination across different tones (values in Hz).

Les quelques passages de yorùbá que nous avons pu entendre suggèrent que cette stylisation rend pour l'essentiel justice à la mélodie des phrases concernées : la durée des syllabes est comparable[1]. En naxi, en revanche, la longueur des syllabes varie fortement : sur la figure 2.25, qui présente les valeurs de F_0 en fonction du temps, les syllabes 1 et 4 se singularisent par leur longueur, plus que par leur mouvement mélodique tel qu'il peut être résumé par des mesures en deux points. Ces syllabes correspondent à des frontières de groupements intonatifs ; la progression de l'énoncé est, groupe par groupe : « moi $_{\text{1e syllabe}}$, à l'intérieur $_{\text{4e syllabe}}$, (je) suis en train de (me) glisser ». Une stylisation en deux points des courbes observées en naxi a néanmoins été tentée (figures à numéro pair, de 2.14 à 2.28), afin de mettre en regard les résultats sur le naxi avec ceux publiés par Laniran et Clements 2003. Il a été choisi de relever les points de données situés à 20% et 90% de la durée totale de la rime : la marge laissée au début est plus grande afin d'exclure la partie de la rime qui est la plus nettement influencée par l'articulation consonantique qui précède ; en l'absence de consonne finale, une mesure du 2^{e} point à 90% a paru raisonnable[2] ; celle-ci soulève en

[1] Les enregistrements en question sont ceux d'Annie Rialland, que nous tenons également à remercier tout particulièrement pour d'utiles discussions.

[2] Le choix d'écarter une certaine proportion de la syllabe tient au souhait de répliquer l'expérience de Laniran et Clements ; d'un point de vue phonétique, il peut paraître plus satisfaisant d'exclure *les n premiers et les n derniers points de données* pour chaque courbe, plutôt qu'une proportion (en %) qui peut ne pas coïncider avec

revanche une difficulté pour les tons montants /LH/, qui comportent fréquemment une descente finale qui s'étend sur plus de 10% de la durée de la rime. Les figures stylisées ne sauraient donc remplacer les courbes moyennes (figures à numéro impair, de 2.13 à 2.27 ; rappelons que la procédure employée pour moyenner les courbes est décrite dans la section 4.4 de l'Annexe 2)[1].

La déclinaison à l'échelle de l'énoncé est estimée de façon approximative en comparant la moyenne de F_0 de la dernière syllabe à celle de la première syllabe. L'abaissement final est relevé séparément, le définissant comme la différence entre la F_0 moyenne de la dernière syllabe et celle des autres syllabes à l'intérieur du même groupement intonatif[2]. Plutôt que de tenter une quantification, cet abaissement a été décrit, de façon qualitative, comme un abaissement de la plage de fréquence fondamentale, une décroissance de fréquence fondamentale au cours de la dernière syllabe, une combinaison des deux, ou une absence d'abaissement final.

L'enregistrement électroglottographique est utilisé pour relever la qualité de voix liée aux frontières (par observation des valeurs de quotient ouvert et inspection du signal électroglottographique). En l'absence de consignes métalinguistiques du type « faites une pause après tel mot », les locuteurs réalisent ou non ces frontières en fonction du débit adopté.

3.2. Résultats : estimation de la déclinaison, et de l'effet des frontières

Le tableau 2.15 indique les principales caractéristiques phonétiques par lesquelles se réalisent déclinaison et frontières dans les énoncés naxi tout au même ton. La signification des abréviations employées dans le tableau 2.15 (ci-dessous) est indiquée dans la légende du tableau. Le terme « groupe » est employé comme un terme générique, sans distinguer mot prosodique, groupe prosodique et groupe mélodique (pour plus de détails, voir l'entrée « Groupement intonatif » du Glossaire).

---

la division de la courbe en cycles glottiques. C'est ce qui est réalisé au chapitre III, dans le calcul de la pente de fréquence fondamentale des syllabes isolées naxi, anglaises et vietnamiennes.

[1] Lorsque des points de données manquent sur toutes les répétitions de la syllabe concernée, la courbe moyenne de quotient ouvert est incomplète, par exemple la courbe de la première syllabe sur la figure 2.23.

[2] Nous ne nous prononçons pas ici sur la désignation à employer pour les groupements intermédiaires entre le mot et l'énoncé, tels que « mot intonatif », « syntagme intonatif », « groupe intonatif », et employons le terme « groupement intonatif » dans un sens large, comme hyperonyme pour ces divers types de groupements. Voir l'entrée **groupement intonatif** du Glossaire.

Ton haut (H)	F2	M4	M5	M7[1]	M9	moy.
Déclinaison à l'échelle de l'énoncé : approximation en tons musicaux	1	1	2	(3)	2.5	1.6
Qualité de voix en frontière de groupe (en fin de 3e syllabe)	R	G	G	(G)	T	
Abaissement final	AD	P	D	(A)	D	
Qualité de voix en fin d'énoncé	R	V	G	(G)	R	

Ton moyen (M)	F2	M4	M5	M7	M9	moy.
Déclinaison à l'échelle de l'énoncé : approximation en tons musicaux	1	1.5	1.5	2	2	1.6
Qualité de voix en frontière de groupe (en fin de 3e syllabe)	R	T	R	R	R	
Abaissement final	D	P	D	D	D	
Qualité de voix en fin d'énoncé	R	V	R	R	R	

Ton bas (L)	F2	M4	M5	M7	M9	moy.
Déclinaison à l'échelle de l'énoncé : approximation en tons musicaux		2	1.5	2	1.5	1.75
Qualité de voix en frontière de groupe (en fin de 3e syllabe)	R	G	G	R	T	
Abaissement final	A	D	D	D	AD	
Qualité de voix en fin d'énoncé	G	G	R	R	R	

Ton montant (LH)	F2	M4	M5	M7	M9	moy.
Déclinaison à l'échelle de l'énoncé : approximation en tons musicaux	1	0	3	4	2.5	2.1
Qualité de voix en frontière de groupe (en fin de 3e syllabe)	R	G	G	R	T	
Abaissement final	P	P	A	D	A	
Qualité de voix en fin d'énoncé	R	G	G	R	T	

Tableau 2.15. Observations sur les principales caractéristiques par lesquelles se réalisent la déclinaison et les frontières des énoncés dont toutes les syllabes portent le même ton. La signification des lettres est la suivante : A pour Abaissement*, c'est-à-dire une plage de F_0 globalement plus basse ; D pour* Descente *(décroissance de F_0 au cours de la syllabe) ; AD pour la combinaison de A et D ; G pour* Glottalisé *; P pour : pas d'abaissement final ; R pour* Relâchement *(remontée du quotient ouvert) ; T pour* Tendu *(absence de remontée du quotient ouvert) ; V pour : variable d'un cas à l'autre. Une case blanche indique que la mesure n'est pas applicable du fait d'une glottalisation.*

Le graphique 2.6 (page 491) reprend les informations quantifiées sur la déclinaison ; il montre une forte variabilité, qui nous paraît être un signe supplémentaire du fait qu'en naxi, même dans le *phonostyle* très contrôlé que représente l'énonciation de phrases préparées à l'avance, les phénomènes locaux (en particulier le marquage des frontières) ont une importance telle que la ligne de déclinaison ne peut en être abstraite.

[1] Les parenthèses signalent que ces résultats ne sont pas sur le même plan que ceux des autres locuteurs : dans le dialecte du locuteur M7 (NL), l'énoncé n'est pas porteur de la séquence tonale recherchée, mais d'une séquence H.H.LH.M.H.H.H.H, du fait que le conditionnel/topicalisateur est prononcé /sè´/, non /´ sé/. Le locuteur a en outre remplacé le pronom /ŋý/ par /nʊ̀/, qui donne /nʊ̄/ lorsqu'il est précédé de la particule marquant l'objet, /tó/ (~ /kó/). Ces différences tonales ne sont pas sans influence sur la baisse de fréquence fondamentale au cours de l'énoncé ; cet énoncé n'a donc pas été intégré aux moyennes.

Pour résumer brièvement les observations : lorsqu'une frontière est réalisée, les syllabes d'un même groupe intonatif sont réalisées à une hauteur similaire (avec seulement une légère diminution d'une syllabe à la suivante, et *modulo* les tons lexicaux), et la dernière connaît (dans les cas typiques) un allongement et un abaissement final. L'allongement final est moindre que l'allongement en frontière de groupe (ce qui rejoint des observations déjà anciennes, par exemple celles de Vaissière 1977 en français). La syllabe qui suit la frontière part typiquement d'une hauteur légèrement rehaussée, phénomène de *réinitialisation partielle de la ligne de déclinaison*.

L'élocution soignée du locuteur M9, à coloration didactique, illustre de façon exemplaire la différence de réalisation en position finale (de groupe ou d'énoncé) : les syllabes y sont environ deux fois plus longues qu'ailleurs (en revanche, chez ce locuteur, la hauteur est tenue : l'abaissement est faible).

3.3. Premières remarques en vue d'une comparaison de la déclinaison entre langues

Yetunde Laniran relate, au sujet du yorùbá :

> Sequences of like tones exhibit a gradual fall in F_0 which is dependent on tone type : L tones show more declination (drop in F_0) than M tones, and H tones show little or no declination. Final tones are subject to a more abrupt lowering at the end of a sentence or phrase. (Laniran 1992:17)

Les résultats du locuteur yorùbá FA montrent néanmoins une déclinaison comparable au ton H et aux deux autres tons (voir la figure reprise ci-dessus, section 3.1.3). En naxi, la déclinaison au ton H est comparable à celle aux tons M et L. Le locuteur M4, chez qui la déclinaison est faible aux quatre tons, ne présente pas plus de déclinaison au ton bas qu'au ton haut. Les figures montrent que la décroissance de F_0 ne se limite pas à un abaissement final.

Bilan au sujet des énoncés dont toutes les syllabes portent le même ton

Sur un plan technique, une mesure en deux points de données ne paraît pas suffisante pour résumer la courbe de F_0 des syllabes naxi : la rime de la dernière syllabe dans chaque groupement intonatif connaît un allongement, souvent accompagné de modifications de la courbe de F_0. Les tons naxi, qui au plan phonologique sont des tons ponctuels, connaissent des variations intonatives, dont le détail phonétique évoque les tons complexes de langues telles que le chinois mandarin (et de nombreuses autres langues, notamment austroasiatiques et hmông-miên). Du fait de cette variation, une modélisation par des courbes paraboliques (telle qu'elle est pratiquée par Andruski et Costello 2004 et Vu-Ngoc Tuân, d'Alessandro *et al.* 2005) paraît plus prometteuse qu'une modélisation par des droites.

Cette caractéristique des tons naxi tient une place importante dans le volet typologique de la Discussion (chapitre IV) : les tons naxi sont des tons ponctuels, mais le détail phonétique de leur réalisation est complexe et ne peut être déduit de la séquence tonale de l'énoncé.

Bilan général au sujet de la composante syntaxique de l'intonation en naxi

Les observations peuvent se résumer de la façon suivante : **la déclinaison n'est pas dissociable des phénomènes de frontières intonatives** (autrement dit, des *morphèmes intono-syntaxiques*); **un patron légèrement descendant s'observe (*modulo* les tons lexicaux) à l'intérieur de chacune des unités intonatives, s'achevant, en fin d'unité, par un allongement de la rime de la dernière syllabe et un abaissement de F_0.**

Table des matières du second volet du chapitre II (« Aperçu d'ensemble de l'intonation du naxi »)

1. Moyens grammaticaux de structuration de l'information ____ 86

1.1. Mise en valeur et mise en retrait ____ 86

1.2. Le rôle de l'ordre des mots ____ 88

1.3. Le rôle des particules grammaticales ____ 88

2. Réalisation d'une intensification par des modifications intonatives locales ____ 89

2.1. Premières observations sur la malléabilité des tons ____ 89

2.2. Le conflit entre ton lexical et mise en valeur intonative ____ 90

2.2.1. Première stratégie de mise en valeur : l'intonation l'emporte sur le ton ____ 91

2.2.2. Deuxième stratégie de mise en valeur (observé seulement sur des voyelles d'arrière) : serrage pharyngien 92

2.2.3. Troisième statégie de mise en valeur : effort supplémentaire qui n'est pas signalé uniquement par un changement de F_0. Choix de la placer au centre de l'expérience principale (rapportée au chapitre III) ____ 92

2.3. Réflexions sur la notion d'*accent d'insistance* ____ 94

2.3.1. Caractérisation de l'accent d'insistance en termes de quotient ouvert : il n'est pas accompagné par une adduction forte des plis vocaux ____ 94

2.3.2. L'accent d'insistance, universel de l'intonation ? ____ 95

2.4. L'ajout d'un ton H intensif, processus catégoriel ____ 98

2.4.1. Description générale du processus ____ 99

2.4.2. L'ajout d'un ton H intensif est distinct de la focalisation contrastive (absente en naxi) ____ 100

2.4.3. L'ajout de ton H intensif est distinct de l'accent d'insistance ____ 101

2.4.4. Perspective typologique au sujet du ton H intensif ____ 102

2.5. Bilan des expériences sur la composante pragmatique de l'intonation en naxi ____ 103

Tableau de l'intonation du naxi, deuxième volet : Recherches sur l'intonation pragmatique de la langue naxi[1]

L'enjeu de ce chapitre est d'évaluer le rôle que les modifications intonatives locales jouent dans la structuration de l'information en naxi. Il s'agit d'établir que le naxi possède des morphèmes intonatifs sémantico-pragmatiques, et de décrire les relations que ceux-ci entretiennent avec les autres moyens offerts par la langue naxi pour la structuration de l'information.

L'exploration de l'intonation pragmatique de la langue naxi proposée ici repose sur une expérience-pilote dont le point de départ est un corpus conçu pour l'étude de la structure de l'information en français, corpus lui-même en partie calqué sur des énoncés anglais (voir Delais-Roussarie, Rialland *et al.* 2002 et références citées). Ce choix fournit l'occasion d'une **comparaison aussi directe que possible entre langues, afin de contribuer à l'éclaircissement des perspectives typologiques**.

Ce corpus est composé de paires questions/réponses ; le terme de *focus* est employé pour désigner *ce sur quoi porte la question*, et *ce qu'apporte la réponse*, orientation terminologique également choisie par Paris 1999:204 et 214, et Xu Yi 1997, 1999. Sans nous dérober à l'exigence de définition des termes employés, précisons que la démarche consiste d'abord à établir des distinctions de fait ; les équivalences de vocabulaire d'un auteur à l'autre sont indiquées pour que le lecteur puisse, le cas échéant, reformuler nos observations dans le cadre terminologique qui lui paraît le plus éclairant.

Les cent trente énoncés du corpus de départ ont pour la plupart subi des modifications pour le propos de la présente enquête. Cette réflexion préliminaire a conduit à l'élimination d'énoncés qui paraissaient trop artificiels, ou non transposables culturellement[2]. Il a été

[1] L'étude présentée dans ce chapitre est une version étendue d'une publication dans les Mémoires de la Société de linguistique de Paris : Michaud 2003a.

[2] Ainsi, en français, le corpus comportait le fameux échange « Qui a tué César ? – Brutus a tué César », calqué sur l'anglais (Chomsky 1972:97). Cette traduction de *Who killed Caesar? – Brutus killed Caesar* n'est en fait pas acceptable, le français ne présentant pas la même latitude que l'anglais dans l'emploi de la topicalisation contrastive (« accent contrastif »). Le souhait d'instaurer un dialogue avec d'autres chercheurs encourage certes à s'inspirer dans la mesure du possible des données linguistiques couramment discutées ; telle est précisément la raison pour laquelle nous avons adopté pour point de départ un corpus existant ; néanmoins, la traduction d'une langue à une autre demande certaines précautions ; il est regrettable de négliger les liens étroits entre stylistique et linguistique, comme c'est le cas dans beaucoup de travaux de syntaxe qui, sur la base de phrases construites de toutes pièces, distribuent les mentions d'agrammaticalité sans réflexion sur les contextes

donné un contexte aux énoncés, pour stabiliser leur interprétation. La langue de l'enquête étant le chinois, nous avons d'abord préparé, avec l'aide d'un locuteur natif du chinois, une traduction respectueuse des habitudes du mandarin familier contemporain (« langue commune » de la Chine). La recherche d'équivalents convaincants de ces phrases en naxi a été menée avec le consultant M1 la veille de l'enregistrement. **La compréhension fine de la structuration linguistique de l'information dans cette langue demande une attention particulière du fait du contexte actuel de bilinguisme naxi-chinois** : au cours d'une enquête de terrain menée à partir du chinois, les calques de structure peuvent masquer certains traits du naxi. Ces traductions ont été notées en API par l'enquêteur ; au cours de la séance d'enregistrement, ce sont les énoncés chinois qui ont été lus au même consultant (M1), qui les a retraduits en naxi l'un après l'autre sans l'aide de documents écrits. Un signal électroglottographique a été enregistré simultanément avec l'enregistrement audio.

1. Moyens grammaticaux de structuration de l'information

Il serait réducteur d'étudier le rôle pragmatique de l'intonation sans mentionner le rôle des moyens grammaticaux offerts par la langue dans l'expression de la structure de l'information. Le naxi ne connaît pas de clivées, à la différence de langues comme le français, l'anglais et le chinois (Paris 1979), mais présente d'autres moyens morphosyntaxiques de structuration de l'information. **Les particules et l'ordre des mots jouent un rôle de premier plan dans les langues tibéto-birmanes**, comme en témoignent notamment les études de Balthasar Bickel sur le bélharien (belhare) (voir Bickel 2003, « Referential density in discourse and syntactic typology », et références citées) et de Martine Mazaudon sur le tamang (Mazaudon 2003 : « From discourse to grammar in Tamang: topic, focus, intensifiers and subordination »).

1.1. Mise en valeur et mise en retrait

Le premier moyen de mise en valeur est négatif : **ne pas expliciter certains syntagmes, ce qui met en valeur par contraste ceux qui demeurent**. L'ellipse du sujet en naxi illustre le paramètre *pro-drop* de la grammaire générative, étiquette glossocentrique que Michel Launey propose de renverser en paramètre « NP-Add » au vu du fait qu'« il est probable qu'au bout du compte le type *pro-drop* apparaîtra largement dominant » (Launey 1994:42-47, 93). Le naxi est, sur ce point, similaire au tamang, autre langue tibéto-birmane, qui connaît une ellipse du « référent actif » (ou « topic »), phénomène décrit dans Mazaudon 2003, § 3.1. Voici un exemple en naxi :

---

d'apparition qui rendraient pleinement licites certaines des phrases préjugées « mal formées » (remarque notamment formulée par Ivan Fónagy [Fónagy 1989]).

(13)	nɯ̀	ə̀ tsɯ̄	bē	nɯ̀	lē?
	2e sg.	pron. interrogatif	faire	part. asp. « en cours »	part. interrog.

« Qu'est-ce que tu es en train de faire ? »

(14)	hā	$t^hý$	nɯ̀.
	nourriture	cuisiner	part. asp. « en cours »

« [Je] fais la cuisine. » (voir également les exemples fournis par He et Jiang 1985:100-101)

En outre, parmi les syntagmes explicités, **la mise en avant de tel ou tel élément d'un énoncé est solidaire de la mise en retrait d'autres éléments**. Dans de nombreuses langues, le post-focus (ensemble des syllabes qui suivent le mot ayant statut de focus) est affaibli : au plan de F_0, de l'intensité, des mouvements articulatoires, cette partie de l'énoncé est peu saillante, ce qui, par contraste, met en valeur le focus (phénomène rapporté pour de nombreuses langues ; voir par exemple Hirst et Di Cristo 1998). L'affaiblissement du post-focus se traduit, au plan morphosyntaxique, par des phénomènes de contraction et d'ellipse. L'exemple suivant illustre la réduction d'une syllabe grammaticale (non focale) lorsqu'un degré élevé d'insistance est attaché à l'élément en position focale.

(15) Tu es venu par le train de 6 heures ?

(16) Non, je suis venu par celui de 7 heures./ (16') Non, par celui de 7 heures.

En français, (16') paraît plus plausible que (16) dans ce contexte. Lors de l'enregistrement de notre corpus naxi, le consultant a prononcé la forme complète (17) ; devant ensuite répéter avec pour consigne « tu l'as déjà dit plusieurs fois ; tu es agacé, impatient », il a répété deux fois (17) puis est spontanément passé à (17') :

(17)	mɤ̄	wà,	ŋɤ̀	ʂə̄	ʤɻɥ̀	gɤ̄	ʈʂʰɯ̄	tjá
NEG.	copule	1e sg.	sept	heure	GEN.	DEICT.	CL.	

ndzɯ̀	pú	tsʰɯ̀	mɤ̀.
être assis	traverserarriver		particule affirmative

« Non, je suis venu par le train de sept heures. »

(17')	mɤ̄	wà,	ŋɤ̀	ʂə̄	ʤɻɥ̀	gɤ̄	ʈʂʰɯ̄	tjá
NEG.	copule	1e sg.	sept	heure	GEN.	DEICT.	CL.	

ndzɯ́	tsʰɯ̀	mɤ̀.
être assis + ton H flottant	arriver	particule affirmative

(même signification. Enregistrement sonore : 17prime_A.wav ; enregistrement électroglottographique : 17prime_E.wav.)

Dans (17'), la particule /pú/ n'apparaît plus : il n'en reste plus qu'une trace tonale, dans le passage au ton montant du ton bas qui précède. Ce phénomène n'est pas conditionné par le contexte phonologique : la réduction de la syllabe /pú/ à un simple ton accompagne le degré élevé d'insistance placé sur un autre élément de l'énoncé, indice d'une solidarité entre mise en valeur de certains éléments et mise en retrait des autres.

1.2. Le rôle de l'ordre des mots

Comme il a été mentionné en tête de ce chapitre, l'ordre des mots joue également un rôle important dans la structuration de l'information en naxi. Dans le domaine sino-tibétain, le rôle de l'ordre des mots est également rapporté pour le tamang par Martine Mazaudon (Mazaudon 2003), par Balthasar Bickel pour le bélharien (belhare ; voir Bickel 1995, 2003), pour le chinois par Marie-Claude Paris (Paris 1999, « Ordre des mots, topique et focus en chinois contemporain »). En naxi, le fait ressort de l'exemple (18), équivalent de la question « C'est *(prénom masculin)* qui est sorti avec *(prénom féminin)* ? »

(18)	ū lí	ū lî	kó	sū sū	lá ?
	nom propre	nom propre	particule objet	avoir une liaison	part. interrogative

C'est nécessairement le premier des deux termes qui est mis en valeur : « C'est /ū li ˊ/ qui est sorti avec /ū li ˋ/ ? » **Dans un cas comme celui-ci, l'intonation à elle seule ne peut changer le sens de l'énoncé** : pour changer de focus, il faut changer l'ordre des mots.

1.3. Le rôle des particules grammaticales

En naxi, plusieurs particules jouent un rôle de premier plan dans la structuration de l'information. La particule /ˊ sé/[1] marque la topicalisation[2], et le conditionnel (dualité qui rappelle les observations de John Haiman : « Les conditionnels sont des topiques » [Haiman 1978]). Un rôle similaire est joué par /t^{h}ū̃/, particule homophone avec le pronom de 3^{e} personne, et avec le *déictique des objets distants*, dont elle provient sans doute (voir He et Jiang 1985:74-75, 88). Une particule /dàˊ/ sert à marquer le *topique contrastif*. Soit l'échange suivant, dans lequel la réplique est en décalage avec la question : « Tu as vu *(prénom masculin)* ? - J'ai vu son blouson dans l'entrée. » En naxi, la réponse a été traduite par :

(19)	t^{h}ū	gɤ̃	bā là	dàˊ	k^{h}ū k^{h}ú	t^{h}è
	3^{e} sg.	GEN.	vêtement	part. de topic contrastif	entrée	GERONDIF

há	jɤ̃.
prédicat d'existence	particule finale

que l'on peut paraphraser par : « son vêtement, lui, se trouve dans l'entrée. » L'emploi de cette particule n'est pas exclusif d'un marquage intonatif : le mot « vêtement », /bā là/, porte dans cet exemple des marques intonatives de focalisation du type décrit plus bas.

Enfin, il existe une particule qui marque l'agent : /nū̃/, et deux qui marquent le patient : /tó/, et /kó/. L'ajout des particules n'est indispensable que lorsqu'il y a inversion sujet-objet

---

[1] Cette particule, historiquement issue d'un ancien /lú sé/, est devenue monosyllabique mais a conservé le ton de la syllabe /lú/ disparue ; celui-ci est devenu un « ton flottant », qui en parole continue se dépose sur la syllabe qui précède, changeant un ton L ou M en contour /LH/ ou /MH/, respectivement ; un ton haut n'est pas affecté. Le diacritique ˊ qui précède la syllabe /sé/ représente ce ton flottant.

[2] Au sujet du terme de « topicalisation », voir Mazaudon 2003 et références citées.

(d'ordinaire, c'est l'agent qui vient en premier). Ces particules étant facultatives dans la plupart des cas, leur emploi, qui explicite le statut actanciel, constitue une forme de mise en valeur de l'argument concerné. **L'inventaire des particules est néanmoins plus limité qu'en tamang. En particulier, il n'existe pas en naxi de particule qui marque spécifiquement le focus**.

2. Réalisation d'une intensification par des modifications intonatives locales

2.1. Premières observations sur la malléabilité des tons

L'enquête vise à établir que les modifications intonatives locales jouent un rôle dans la structuration de l'information ; autrement dit, que le naxi possède des morphèmes intonatifs sémantico-pragmatiques. Cela est apparu de façon manifeste dès le stade de l'enquête de vocabulaire : les mots étaient dits par les consultants (M1 et F1) dans une phrase-cadre, « je dis __ » (en naxi : /ŋɤ̀ ... ʂɤ́/ : 1^{e} sg.-*syllabe cible*-dire) ; le comportement de la phrase-cadre donne une idée de la souplesse tolérée par les tons lexicaux, et de son exploitation pragmatique. Entre le cadre, qui ne varie pas, et la syllabe encadrée, la hiérarchie au plan informatif est claire ; elle se traduit phonétiquement par une saillance de la syllabe encadrée, solidaire d'une mise en retrait des syllabes du cadre. Le phénomène est particulièrement aisé à observer lorsque la syllabe-cible porte, comme le verbe /ʂɤ́/ final, un ton lexical haut. Dans cette configuration, F_0 est de 5% moins élevée en moyenne sur la syllabe finale du cadre, /ʂɤ́/, que sur la syllabe au ton haut, le quotient ouvert est légèrement plus élevé (de 7%), ce qui signale une moindre adduction des plis vocaux. (Les enregistrements correspondants sont syllabesHML.wav, syllabesLMH.wav et syllabes MLH.wav, dans le dossier du locuteur M1.) Il ne s'agit pas d'un cas de *downstep* (au sujet de cette notion, voir l'entrée correspondante du Glossaire), car la fréquence fondamentale du verbe /ʂɤ́/ est plus basse encore après un ton M ou L qu'après un ton H.

Une autre illustration de la variabilité des tons est fournie par les structures dans lesquelles figurent deux syllabes de même composition phonémique et de même ton. Le fait peut paraître anecdotique, mais est révélateur de phénomènes omniprésents dans la prosodie du naxi. Ainsi, « un lac » se dit :

(20)	hʉ́	dʉ̄	hʉ́
	lac/mer	un	*classificateur des lacs et mers*

Le classificateur est issu du nom, dont il demeure homophone. Cette configuration offre l'occasion d'une comparaison directe des paramètres intonatifs. Les figures 2.29 et 2.30 présentent les courbes de F_0 et quotient ouvert d'une réalisation de cette structure (hors énoncé) par le locuteur M1. **Cet exemple illustre les variations prosodiques que peut connaître une même syllabe tonalisée en fonction de son statut grammatical et de son importance dans l'énoncé**. Le classificateur, mot grammatical, a par là une importance

prosodique moindre que le nom (il figure en outre en position finale). Sa fréquence fondamentale est sensiblement plus basse que celle du nom. Le quotient ouvert est plus élevé sur le classificateur, ce qui signale une phonation plus détendue. Les courbes d'intensité et de DECPA manifestent la même tendance. (Au sujet de DECPA, voir l'Annexe 2, section 3.)

2.2. Le conflit entre ton lexical et mise en valeur intonative

Que devient le ton bas sous insistance ? Subit-il une montée de plage de F_0 ou au contraire un rabaissement qui revienne à réaliser superlativement sa cible lexicale basse ? Cette question nous a souvent été posée[1].

En naxi, le ton H possède une marge de liberté supérieure à celle des deux autres tons : il ne correspond pas au sommet de sa tessiture, et n'est contraint par aucun ton plus élevé que lui dont il doive être maintenu distinct. **En revanche, le ton M ne possède qu'une marge de liberté relativement étroite, entre les domaines du ton L et du ton H. Quant au ton L, il est contraint par la plus faible marge de liberté de la voix dans les basses fréquence** : en théorie, le locuteur peut exploiter tout le bas de sa tessiture pour réaliser le ton L, mais la partie de la tessiture au-dessous de la plage moyenne du locuteur est moins étendue que la partie au-dessus, mise à part la latitude d'un passage en mécanisme 0 (c'est-à-dire une *laryngalisation* ; au sujet des mécanismes phonatoires, voir Roubeau, Chevrie-Muller *et al.* 1987, 1991, Henrich, Roubeau *et al.* 2003). Ainsi, parmi les cinq niveaux intonatifs distingués par M. Rossi et M. Chafcouloff sur une base psychoacoustique, trois sont au-dessus du niveau *Médium : Suraigu, Aigu, Infra-aigu*, contre un seul au-dessous *(Grave)* (Rossi et Chafcouloff 1972). S'il est permis de résumer en anglais : ***the H tone has some headroom; the L tone barely has any 'footroom'***. Cette asymétrie se traduit par un écart-type de F_0 plus élevé au ton H qu'au ton M et au ton L, comme le montrent (au chapitre III) les figures 3.35 à 3.49, qui présentent les trois tons de base prononcés en phrase-cadre par une locutrice et trois locuteurs. Cette situation a une incidence sur les possibilités de mise en valeur prosodique d'une syllabe : une tendance universelle veut qu'une F_0 basse signale une *faible importance* prosodique, la *mise en valeur* étant généralement associée à une valeur élevée de F_0 (voir notamment Vaissière 1983, 1995). Les variations que connaît le ton H en naxi rappellent les phénomènes de mise en valeur dans les langues sans tons lexicaux, tels qu'ils sont décrits par Klaus Kohler (Kohler 1991c ; pour plus de détails, voir la section 1 de l'Annexe 3) : la réalisation soignée (telle qu'elle apparaît sur les syllabes lues en cadre, figures 3.35 à 3.49) est plane ; une descente rapide dès le début de la voyelle, partant d'une valeur de F_0 élevée, exprime l'évidence, l'acquis. (Cette dernière observation a été effectuée, de façon répétée, en parole dite « spontanée », en particulier dans des récits ; sa validation statistique n'a pas été tentée, du fait des problèmes que soulève l'application des méthodes

[1] Notamment par les relecteurs d'une communication au colloque « Speech Prosody 2004 », et lors de la préparation de la Journée scientifique de la Société de linguistique de Paris sur la focalisation.

statistiques à des données de ce type.) **Dans le cas des tons M et L, le corpus de l'étude-pilote révèle plusieurs stratégies pour résoudre le conflit entre ton lexical L ou M et mise en valeur prosodique**.

2.2.1. Première stratégie de mise en valeur : l'intonation l'emporte sur le ton

Une stratégie consiste, dans les cas où le contexte limite les risques de confusion, à passer outre au ton lexical : l'intonation l'emporte sur le ton ; la fréquence fondamentale du ton L se trouve alors emportée dans la plage du ton M. En effet, la présence de tons lexicaux ne signifie pas que chacun doive être pleinement réalisé ; les tons sont en cela semblables aux phonèmes, qui (dans toutes les langues) peuvent, selon le contexte prosodique, être renforcés ou affaiblis, modifications parfois extrêmes que les auditeurs parviennent à compenser dans une large mesure[1]. Ainsi de l'exemple sonore (21), dû à la locutrice F1 :

(21) ŋɤ̀ sè bē lɔ̄ lē?
1e sg. *pronom interrogatif* « comment » devoir *particule interrogative*

« Que dois-je faire ? » *(récit « Origine », à 5 mn 36 sec. 500 ms ; extrait sonore correspondant : 21_A.wav)*

Dans ce contexte, un morphème intonatif pragmatique porte sur le pronom de première personne ; le sens est : « Et moi, qu'est-ce que je vais faire ? » La fréquence fondamentale de la rime de la première syllabe se trouve dans la plage du ton M de la locutrice.

Ce phénomène demeure phonétique : lors de la transcription, les consultants sont unanimes pour dicter la syllabe /ŋɤ̀/ (1e pers. sg.) avec un ton L. Mais il paraît pouvoir être relié à un phénomène phonologisé du naxi : la règle de remontée du ton des pronoms personnels de 1e et 2e personne (/ŋɤ̀/, /nù/, tous deux au ton L) lorsqu'ils sont suivis de la particule (optionnelle) /nū/, qui signale le statut d'*agent* ou d'*origine*, ou de la particule /to´/ ou /ko´/, qui signale le *patient* ou le *destinataire*. Ainsi, /ŋɤ̀/ + /nū/ se réalise /ŋɤ̄ nū/, /nù/ + /nū/ se réalise /nū nū/. (Intonation pragmatique et intonation syntaxique paraissent concourir à ce changement tonal ; une hypothèse la situation historique dans lequel la variation allotonique a pu déboucher sur un changement tonal est formulée dans le premier volet du présent chapitre, section 2.2.2.)

Comme le ton L, le ton M peut être emporté dans une plage de fréquence fondamentale élevée par des facteurs intonatifs.

---

[1] Voir par exemple Manuel 1995, « Speakers nasalize /ð/ after /n/, but listeners still hear /ð/ » ; Holt et Kluenderb 2000, « General auditory processes contribute to perceptual accommodation of coarticulation » ; Kohler 1990, 2000, 2001.

2.2.2. Deuxième stratégie de mise en valeur (observé seulement sur des voyelles d'arrière) : serrage pharyngien

Une autre solution pour résoudre le conflit entre ton lexical et mise en valeur prosodique est un serrage pharyngien qui aboutit au dévoisement (partiel ou complet) de la syllabe. Ces exemples se rencontrent assez fréquemment aussi bien en voix féminine (récit de F1) qu'en voix masculine (récits de M4). Ainsi dans un récit de la locutrice F1 :

(22)	$k^hɣ̀$ kỳ kò	à	dɯ̄	mē	ə̀	$k^huú$...
	à l'intérieur	poulet	un	*classificateur*	*particule d'aspect*	mettre

« Dedans, tu mettras un poulet » (enregistrement : à 5 mn 26 sec 106 ms du récit « Origine » ; extrait sonore : 22_A.wav)

A ce point du récit, le mot /à/ (« poulet ») est mis en relief ; **l'enregistrement électroglottographique montre un dévoisement presque complet, le voisement ne s'installant que tout à la fin de la voyelle. Pendant la partie dévoisée, la source sonore n'est pas la vibration des plis vocaux mais paraît être une intense friction pharyngale**, qui signale efficacement la saillance prosodique de cette syllabe, là où un segment voisé aurait difficilement pu porter à la fois la même charge intonative et le ton lexical, nécessaire à la bonne identification de ce mot non prédictible à partir du contexte. Cette stratégie de mise en valeur paraît beaucoup plus commune au ton L qu'au ton M ; néanmoins, elle est possible au ton M, et même au ton H[1].

2.2.3. Troisième statégie de mise en valeur : effort supplémentaire qui n'est pas signalé uniquement par un changement de F_0. Choix de la placer au centre de l'expérience principale (rapportée au chapitre III)

Dans la majorité des cas de mise en valeur locale d'une syllabe au ton bas, la réalisation du morphème intonatif d'insistance ne va pas jusqu'à dévoiser la voyelle, ni rendre le ton méconnaissable.

Un certain nombre d'énoncés donnent l'impression auditive d'un **effort supplémentaire qui n'est pas signalé uniquement par un changement de F_0.** Soit l'exemple suivant :

(23)	nɯ̀	$k^hà$ dỳ	$t^hȳ$	bɯ̄	lá ?
	2e pers. sg.	Pékin	sortir	aller	*particule interrogative*

« Tu vas jusqu'à Pékin ? » *(enregistrement audio :* 23_A.wav *; enregistrement électroglottographique : 23_E.wav)*

Avec mise en valeur intonative sur /$k^hà$ dỳ/ (« Pékin »), la phrase signifie : « C'est bien à Pékin que tu vas ? » Les tracés obtenus sont présentés sur les figures 2.31 et 2.32. La mesure de la mise en valeur intonative est délicate. Ni le quotient ouvert ni F_0 ne présentent

[1] Pour compléter sa description, il serait utile d'étudier sa fréquence relative en fonction de la voyelle : le phénomène paraît moins fréquent dans le cas des voyelles d'avant ouvertes, telles que /i/. Au plan de la description expérimentale, il serait intéressant d'obtenir des images du pharynx.

de saillance particulière sur le mot « Pékin », /kʰà dỳ/. En termes de hauteur, la F_0 sur « Pékin » (/kʰà dỳ/) est plus basse que sur le pronom ; en termes de pente de F_0, « Pékin » est moins saillant que le verbe qui suit, /tʰɣ̄/, ceci conformément à l'opposition lexicale entre les tons L du nom /kʰà dỳ/ et le ton M du verbe /tʰɣ̄/. « Pékin » ne présente pas non plus de saillance en termes de durée. Au plan de l'intensité acoustique, la syllabe la plus intense de l'énoncé est la négation, en début d'énoncé. Seule la mesure de DECPA (mesure expliquée en Annexe 2, section 3) révèle une saillance sur la première syllabe de « Pékin », /kʰà dỳ/. **Cette observation rejoint celles de Klaus Kohler sur ce qu'il nomme « non-pitch accents »** (Kohler 2003).

Il convient de signaler la marge de liberté qui existe dans la réalisation de semblable mise en valeur : le phénomène local sur le mot-focus peut être remplacé par un phénomène global d'élévation de la plage de F_0, la plus forte saillance prosodique se reportant sur la fin de l'énoncé. Telle est la configuration observée dans l'énoncé (24), représenté en figure 2.33. (Pour représenter trois paramètres sur une même figure sans qu'ils se confondent, DECPA a été représenté en valeurs négatives.)

(24) ŋɤ̀ kʰà dỳ bū mɤ̀.
1e sg. Pékin aller *particule finale d'affirmation*

« C'est à Pékin que je vais. »

Les paramètres relevés suggèrent (en accord avec l'impression auditive) qu'il n'y a pas de phénomène local de mise en valeur du mot « Pékin », /kʰà dỳ/, le quotient ouvert indiquant une phonation relativement détendue. Le ton lexical L de la particule finale est méconnaissable : sa courbe de F_0 domine celle de tout l'énoncé ; ses caractéristiques d'intensité et de DECPA sont pareillement saillantes.

Comme le ton L, le ton M peut porter trace d'un effort supplémentaire qui ne bouleverse pas pour autant la courbe de F_0. Ainsi dans un échange qui, très classiquement, vise à mettre en valeur un des éléments de la réponse (déjà citée ci-dessus, sous le numéro 17') :

(25) nɯ̀ hù mū kʰɣ́ ʈʂʰwá ɖʐɻ̀ gɤ̄ hō ʈʂʰɤ̄ ndzɯ̀ pú tsʰɯ̀ lɑ́?
2e sg. soir six heure *GEN.* train asseoir *gérondif*arriver *part. interrog.*

« Tu es venu par le train de 6h du soir ? » *(enregistrement audio :* 25_A.wav *; enregistrement électroglottographique : 25_E.wav)*

(17') mɤ̄ wà, ŋɤ̀ ʂə̄ ɖʐɻ̀ gɤ̄ ʈʂʰū kjɑ́ ndzɯ̋ tsʰɯ̀ mɤ̀.
NEG. *1e sg.* sept heure *GEN.* *déict.* *CL.* asseoir + *gérondif* arriver *part.*

« Non, c'est par celui de 7 heures que je suis arrivé. » *(enregistrement audio :* 17prime_A.wav *; enregistrement électroglottographique : 17prime_E.wav)*

De la figure 2.34, il ressort que la fréquence fondamentale de la syllabe « sept », /ʂə̄/, est inférieure à celle du classificateur /kjɑ́/. La figure 2.35 révèle une brève saillance de DECPA sur la syllabe-focus « sept », indice, comme dans l'exemple au ton bas, d'un phénomène local de mise en valeur prosodique, qui n'est pas décelable sur la seule courbe de F_0, ni sur

la courbe d'intensité globale, sur laquelle le focus ne se détache pas plus que les deux premières syllabes de l'énoncé.

L'existence de ce type de modification motive et légitime l'expérience menée au chapitre III. Pour anticiper quelque peu sur ses résultats : les hypothèses formulées ici y trouvent confirmation ; au plan technique, l'enregistrement des mêmes mots sous deux conditions de lecture autorise la comparaison de leur intensité acoustique globale, mesure qui n'est pas interprétable dans l'expérience-pilote (du fait de la variété des phonèmes et de la position des mots concernés à l'intérieur de l'énoncé), où est utilisé le paramètre DECPA, qui vise à offrir un indice d'effort vocal indépendant des phonèmes, mais présente des limites (détaillées dans l'Annexe 2, section 3).

2.3. Réflexions sur la notion d'*accent d'insistance*

2.3.1. Caractérisation de l'accent d'insistance en termes de quotient ouvert : il n'est pas accompagné par une adduction forte des plis vocaux

Le quotient ouvert fournit une indication sur le degré d'adduction des plis vocaux, et par là sur le caractère plus ou moins *pressé* de la voix. **La mesure de quotient ouvert était d'abord guidée par l'espoir d'utiliser ce paramètre comme indice de mise en valeur prosodique**, escomptant une valeur de quotient ouvert basse sur les syllabes mises en valeur prosodiquement. Cette hypothèse était notamment suggérée par une observation de M. Epstein sur l'anglais : « Both prominent words and phrase-initial words displayed a 'tenser' voice quality than their non-prominent and phrase-final counterparts » (Epstein 2002:xviii, *Voice Quality and Prosody in English*). Sur l'ensemble du corpus, cette recherche n'a pas abouti au résultat escompté. Au contraire, dans les énoncés prononcés avec le degré le plus élevé d'insistance (degré évalué de façon subjective à l'écoute), les syllabes focalisées présentent des valeurs de quotient ouvert élevées. Ainsi, dans les trois réalisations de l'énoncé (17') obtenues en demandant explicitement un degré élevé d'implication du locuteur (degré élevé d'*impatience*), la syllabe /ʂə̄/ (« sept ») présente un quotient ouvert de 56% en moyenne, donc une vibration moins *pressée* que sur les deux premières réalisations (moyenne : 47%), qui ne portaient pas la même charge d'insistance et d'émotion (ces deux dimensions n'ont pas été séparées nettement).

Deux résultats peuvent être retenus :
- le quotient ouvert (à lui seul) ne peut être interprété directement comme indice de focalisation/mise en valeur prosodique
- le quotient ouvert est relativement élevé dans certains exemples qui se rattachent à la catégorie de l'accent d'insistance.

Ce dernier résultat va à l'encontre d'observations sur la voix chantée, qui montrent, chez un chanteur qui élève l'intensité tout en tenant la même note, **une diminution du quotient ouvert**

proportionnelle à l'augmentation de l'intensité (figure 8.7.B3 de Henrich 2001). Pour comprendre ce qui a lieu dans les données de voix parlée dont il est ici question, il paraît éclairant de mettre en parallèle les mesures apportées par l'électroglottographie avec les données qui ont été publiées, pour de nombreuses langues, au sujet de l'*accent d'insistance*.

2.3.2. L'accent d'insistance, universel de l'intonation ?

L'accent d'insistance, en tant que mécanisme physiologique, paraît être un universel, qu'il importe de distinguer des phénomènes de focalisation (et des phénomènes d'*accentuation libre* et d'*accentuation fixe* ; le terme « *accent* d'insistance » perpétue le risque de confusion entre ces phénomènes, mais demeure utilisé ici pour éviter de s'écarter de l'emploi courant). Les premières descriptions de l'accent d'insistance ont été un temps oubliées, du fait de la prééminence accordée à F_0, paramètre aisé à mesurer, et à manipuler en synthèse de la parole. L'accent d'insistance, qui se prête à une définition en termes physiologiques, est pourtant une composante essentielle de l'intonation (voir par exemple Kohler 2003:2 : « ...it is necessary to complement the accent-pitch link by a force accent category [...] It is related to greater physiological and articulatory effort [...] »).

Les descriptions de ce phénomène se font écho d'une langue à l'autre. En français, les conclusions de l'enquête physiologique et perceptive de Carton, Hirst, Marchal et Séguinot (Carton, Hirst *et al.* 1976) sont résumées ainsi par Fernand Carton (c'est nous qui avons mis en gras un passage) :

> Il n'y a pas de réalisation constante au niveau acoustique. C'est ce qui nous a amené à chercher s'il n'y a pas une réalisation physiologique constante. (...) On remarque que **les accents d'insistance qui ont été repérés par un très fort pourcentage d'auditeurs sont ceux où apparaît une *brutale montée* de pression d'air sous-glottique qui commence sur une *consonne*.** (...) Ce qui est caractéristique, c'est la brusquerie inattendue de l'effort expiratoire. Dans le cas de l'accent normal de groupe, quand l'accent est peu expressif, la montée de la courbe de pression sous-glottique est lente. Que la valeur soit « affective » ou « intellectuelle », l'énergie expiratoire se développe avec une pareille soudaineté. Elle s'associe dans les deux cas à l'énergie articulatoire, mais celle-ci est beaucoup plus difficilement mesurable directement (la durée en est une mesure indirecte). (...) Dans la grande majorité des cas, il y a accroissement parallèle de la pression d'air et de la hauteur des sons émis. Mais la fréquence vibratoire des cordes vocales est réglée aussi par la tension musculaire du larynx, dans des conditions encore mal connues. C'est pourquoi la hauteur monte sur certains tracés alors que la pression baisse, ou inversement. (Carton 1974:121)

Un dispositif expérimental recourant lui aussi à une technique invasive de mesure de pression d'air sous-glottique a été mis en place par André-Pierre Benguerel, qui conclut, reprenant l'opposition entre *accent emphatique* et *accent non emphatique* proposée par Hélène Coustenoble et Lilias Armstrong (Coustenoble et Armstrong 1937) : « ... dans le cas de l'accent emphatique, il y a toujours une augmentation marquée de la pression sous-

glottique pendant (ou juste avant) la syllabe accentuée » (Benguerel 1973, « Corrélats physiologiques de l'accent en français »). Ce prototype physiologique de l'accent d'insistance ressort également d'une étude pléthysmographique (mesure du volume pulmonique), qui montre, de façon invariable, une diminution du volume pulmonique (indiquant un surcroît d'activité expiratoire) porteuses d'une insistance (« ... increase in expiratory muscular activity only on emphatically stressed syllables, not on conversationally stressed syllables » : Ohala 1978:21). Une étude de Fant, Hertegard et Kruckenberg, dans laquelle apparaît la formule d'« activité pulmonique prosodiquement significative », distingue également deux types de comportement :

> ... an increased pulmonic (and hence subglottal) pressure is observed in cases of contrastive and higher levels of stress, but does not appear to be a necessary component of focal accentuation. (Fant, Hertegard *et al.* 1996, « Focal accent and subglottal pressure »)

Les dénominations varient ; il paraît néanmoins possible de retenir la convergence de ces observations approfondies. Au plan phonétique, l'accent d'insistance, compris comme un geste d'*effort*, apparaît comme une catégorie universelle[1], définissable en termes physiologiques. Au plan des fonctions linguistiques, il peut remplir des fonctions pragmatiques (manifestation extrême de la *mise en valeur intonative*, et de la *focalisation contrastive* dans les langues qui connaissent ce procédé) ; il a également part (d'une façon difficilement dissociable de son rôle au plan pragmatique) à l'expression des attitudes et émotions. La fréquence d'emploi de l'accent d'insistance peut certes varier selon les systèmes prosodiques ; en français, son emploi systématique dans le débit dit « professionnel » paraît en effet récent. En revanche, le mécanisme phonétique de l'accent d'insistance est sans aucun doute présent dans toutes les langues (avec un poids phonostylistique inversement proportionnel à sa fréquence). Ainsi, le wolof paraît posséder une forme d'accent d'insistance, certes peu fréquent (Rialland et Robert 2001, 2003) ; dans l'interprétation qu'en proposent ces auteurs, le phénomène en question est transcrit sous forme de *tons*, suivant en cela le courant autosegmental-métrique d'études intonatives, mais des discussions avec Annie Rialland suggèrent que le mécanisme physiologique pourrait être du même type que dans une langue comme le français (étant entendu que des relations d'équivalence perceptive peuvent exister entre l'augmentation de pression sous-glottique, réalisation prototypique de l'accent d'insistance, et d'autres phénomènes, tels que l'allongement des consonnes ou les mouvements de fréquence fondamentale, en vertu de ce que Gussenhoven 2002 nomme le « code de l'effort », *« effort code »* ; voir également Vaissière 2004 et références citées).

---

[1] Il ne paraît donc pas crédible d'affirmer avec Fónagy (1983:22) : « L'allongement consonantique et l'accent d'insistance, si courants dans le français moderne, étaient probablement inconnus avant la deuxième moitié du XX^{e} siècle ».

Pour revenir, au terme de ce panorama de travaux publiés au sujet de l'accent d'insistance, aux énoncés du corpus naxi, les valeurs de F_0 et d'intensité acoustique relevées lors de l'étude-pilote encouragent à désigner les cas extrêmes d'insistance observés comme réalisation d'un *accent d'insistance*. Définir le phénomène physiologiquement, par le surcroît d'action des muscles respiratoires, conduit à le distinguer nettement parmi les phénomènes prosodiques (sans exclure l'existence de réalisations variées, en vertu des relations d'équivalence qui viennent d'être évoquées). **Les données électroglottographiques sur la langue naxi encouragent à ajouter à la caractérisation de l'accent d'insistance (dans sa réalisation la plus typique) l'absence de serrage des plis vocaux**, la voix n'étant pas *pressée* sur la syllabe qui porte l'accent d'insistance. Soulignons la coïncidence entre le quotient ouvert de 57% relevé ci-dessus pour les syllabes portant un accent d'insistance et la valeur relevée par Ingo Titze :

> ... the Oq = 0.5 half-sinusoid is ideal for low pressures, but for high pressures (and higher fundamental frequencies), a slightly abducted vocal fold *(sic.)* gives optimum dbA power. At $P_S = 24$cm H_2O, for example, an abduction quotient of 0.2 (open quotient of 0.56) seems to be optimal. (Titze 1988:236)

Le paramètre de durée entre également en compte ; pour faire ressortir cette variation, une comparaison a été menée entre la syllabe /ʂə̃˞/ « sept » telle qu'elle est réalisée dans (17') et telle qu'elle apparaît dans deux autres contextes :

1. la syllabe /ʂə̃˞/ « sept » dans des listes de nombres associés à un classificateur (un classificateur au ton bas : « montagne », un au ton moyen : « animal », un au ton haut : « arbre » ; les listes vont de 1 à 25)

2. la syllabe homophone /ʂə̃˞/ « chose », dans l'énoncé (26) :

(26)	ə tsɯ̄	gɤ̃˞	ʂə̃˞	$t^hɥ̄$	jɤ̃˞?
	pronom interrogatif	POSS.	chose	se passer	particule finale
	« Qu'est-ce qui se passe ? »				

Il apparaît que le rapport *durée de la consonne/durée de la voyelle* est nettement plus élevé pour les items focalisés (porteurs ou non de l'accent d'insistance)[1], où il est de l'ordre de 1.5, que dans les deux autres contextes, où il est de l'ordre de 0.8. L'importance des durées respectives de l'initiale et de la rime d'une syllabe est connue notamment par les travaux de Vaissière 1977, 1986 et Astesano 1999. Intonation pragmatique et expression de l'émotion paraissent jouer du même mécanisme :

> ... à l'intensité de l'émotion correspond l'intensité de l'activité musculaire, ainsi que le prolongement des voyelles et des consonnes (...) : celui des voyelles pour les

[1] Dans ce calcul, la durée de la consonne est définie comme l'intervalle non voisé entre la fin de la voyelle précédente et le début du voisement de la voyelle /ə˞/. Il n'est pas tenu compte de l'allongement de la rime de la syllabe précédente, qui contribue à la mise en valeur de la syllabe /ʂə̃˞/.

> émotions tendres, celui des consonnes pour les émotions agressives. (Fónagy 1983:18)

Une comparaison directe des rapports consonne/voyelle ne peut se faire qu'entre syllabes identiques au plan phonémique, c'est pourquoi la comparaison se limite ici à la seule syllabe /ʂə̄/ ; des observations non quantifiées tendent à confirmer l'importance de ce paramètre dans le reste du corpus de cette expérience-pilote.

2.4. L'ajout d'un ton H intensif, processus catégoriel

Des observations qui précèdent, il ressort que la mise en valeur intonative des tons L et M pose bel et bien un problème phonétique, résolu par diverses stratégies. **Le naxi possède en outre un mécanisme catégoriel qui a précisément pour fonction de mettre en valeur une syllabe au ton L ou au ton M, par *l'ajout d'un ton H.*** L'exposition de ce phénomène original (même s'il n'est pas sans équivalent dans d'autres langues, d'Asie et d'ailleurs) demande des explications relativement détaillées sur la place du ton montant et des tons flottants dans la langue naxi.

Outre ses tons H, M et L, le naxi occidental possède un ton montant, plus rare, noté /˩˥/ dans les sources chinoises. Il apparaît sur quelques mots « natifs » et sur un nombre croissant d'emprunts au chinois. La présence d'un ton montant sans ton descendant correspondant a quelque chose de surprenant au plan typologique. La situation inverse est fréquente : une langue possédant un contour descendant et pas de contour montant ; par exemple, dans le domaine tibétain, le shigatse et le dzongkha (Sun 1997:487, 502n et références citées). Les contours montants sont évités dans de nombreuses langues (notamment dans la branche bantu de la famille Niger-Congo : voir Kisseberth 2001:149 ; et dans des dialectes japonais : Nakai 2001). Des stratégies d'évitement sont observées, telles que le remplacement d'une montée par un simple ton H (par exemple en hausa, dans la branche chadic de la famille afro-asiatique ; voir une présentation détaillée par Newman 1995:765-766), ou le déplacement de la partie H du contour montant vers la syllabe suivante. Parmi les systèmes à trois tons d'Afrique (tels que ceux observés en gulmancema, ncam, kasem dans la famille voltaïque, et en igbo, yoruba, yala-ikom, ega, ebrie dans la famille kwa ; voir Rialland 1998:409), ceux qui autorisent les contours montants autorisent aussi des contours descendants ; l'un et l'autre type de contour proviennent de l'association de deux tons différents à une même syllabe. Au plan physiologique, un contour descendant est produit plus rapidement, et de façon qui paraît plus aisée, qu'un contour montant (voir Ohala et Ewan 1973 ; Ohala 1978:30-31 ; Sundberg 1979, et diverses études de cas plus récentes, par exemple le travail de Nagano-Madsen 2003 sur le dialecte japonais de Kochi). La situation en naxi attire donc l'attention.

Le ton montant du naxi tend à être considéré comme un élément étranger, du fait qu'il apparaît sur de nombreux mots d'emprunt au chinois (Li Lincan, Zhang Kun *et al.* 1953:57 ; Rock 1963-1972 ; Bradley 1975:95 ; Fang Guoyu et He Zhiwu 1995 ; Pinson 1998:xviii).

Néanmoins, une courte description du naxi formule l'hypothèse selon laquelle il jouait à l'origine un rôle morphologique ou syntaxique[1].

Nos données confirment qu'il existe peu de mots portant un ton montant, une fois mis à part les emprunts au chinois ; elles démontrent en revanche qu'en parole continue, la réduction d'un morphème au ton H peut aboutir à la création d'un contour montant sur la syllabe qui précède : le contenu segmental du morphème disparaît tout à fait ; le ton H *flotte* alors (pour reprendre la terminologie en usage dans la description des tons subsahariens) et se réassocie à la syllabe qui précède. Les morphèmes concernés sont par exemple /tá/ « seulement », /nú/ « ou », et le conditionnel /lú/ (les détails sont présentés dans un article soumis pour publication : « Contour tone emergence, tonal reassociation and intonational contours: evidence from Naxi »).

Un second processus conduisant, en discours, à l'apparition de contours montants est l'ajout d'un ton H à une syllabe portant un ton lexical M ou L ; ce processus d'*intensification intonative* transforme le ton d'origine en un contour montant (MH ou LH, selon le ton lexical d'origine).

2.4.1. Description générale du processus

Dans les dialectes occidentaux du naxi, un mot dont la dernière syllabe porte un ton L ou M peut recevoir une intensification par ajout d'un ton H. Le phénomène est optionnel, au sens où il n'est pas prédictible à partir du contexte phonologique. Le ton H ainsi ajouté s'associe à la dernière syllabe du mot, constituant avec le ton lexical d'origine une séquence LH ou MH, par exemple dans l'énoncé suivant, extrait du récit « Coutumes » (à 9 mn 43 sec sur l'enregistrement) :

(27) …	hā pǎ́	gɤ̃	lā	dū hū	kó.
	chinois	GEN.	aussi	quelques	fêter

« [Les Naxi ne célèbrent pas seulement leurs propres fêtes, ils] célèbrent aussi quelques-unes des fêtes chinoises. »

La forme lexicale du mot « chinois » est /hā pà/. Dans l'énoncé (27), le mot est transformé en /hā pǎ́/ ; l'effet énonciatif est de mettre en relief ce mot, à un point de transition dans le récit, entre description de traditions anciennes perçues comme autochtones et description de coutumes adoptées de la culture chinoise à date récente. Au vu de l'existence en naxi de

[1] La formule est la suivante :
« 声调方面，早期纳西语只有高平，中平，低降三个基本声调，但后来增添了一个低升调。低升调的产生，可能与复元音产生的情况相同。最初这个声调只出现在个别的单词和句子中，起了构词或某种语法作用。/ En ce qui concerne les tons, le naxi ancien ne possédait que trois tons de base, haut-plain, moyen-plain et bas-descendant ; par la suite s'y est ajouté un ton bas-montant. L'apparition du ton bas-montant est peut-être similaire à l'apparition des noyaux vocaliques complexes. A l'origine, ce ton n'apparaissait que sur certains mots isolés et dans certaines phrases ; il devait jouer un rôle de dérivation lexicale, ou un rôle grammatical d'une sorte ou d'une autre » (He Jiren et Jiang Zhuyi 1985:41).

processus d'ellipse de mots grammaticaux aboutissant à la réassociation de leur ton H, la première hypothèse a été qu'il s'agissait d'un cas d'ellipse syllabique. Diverses manipulations ont été tentées pour rétablir une syllabe pleine en lieu et place du ton H, entre le mot /hā pà/ et le possessif /gɤ̄/. Les énoncés manipulés ont été refusés par les consultants ; au plan diachronique, il n'est pas possible d'exclure l'hypothèse selon laquelle le ton H d'intensification proviendrait de la réduction d'une syllabe (un ancien morphème ayant partie liée avec l'intensification ; par exemple un topicalisateur, tel que le conditionnel-topicalisateur /lɯ́/ ; ou la particule /ʂɤ́/, qui joue également le rôle de topicalisateur), mais si tel est le cas, le ton H résultant a depuis lors suivi son propre chemin, car il apparaît dans des positions dans la phrase où aucune des particules aujourd'hui existantes en naxi ne peut lui être substituée. Il n'existe pas de raison impérieuse de postuler que ce ton provient d'une réduction syllabique.

Deux tons H d'intensification (ou plus) peuvent apparaître sur des mots voisins, par exemple dans l'énoncé 28 (extrait d'un dialogue). Conformément aux attentes, l'effet des intensifs qui se succèdent est cumulatif :

(28) pɔ́ sé gjà́ ŋgjɤ́ mɤ̀ !
écrire COND./TOP. très + H intensif pénible + H intensif part. fin. affirmative
« [Apprendre le chinois ça n'est pas trop difficile, mais] l'écriture est vraiment très, très difficile ! »

Le caractère optionnel du ton H intensif ressort clairement de la variabilité des réalisations chez les consultants au cours des séances de transcription : ils omettent parfois par inadvertance un tel ton H lorsqu'ils répètent un passage enregistré, ou au contraire en ajoutent un qui n'est pas présent sur l'enregistrement.

2.4.2. L'ajout d'un ton H intensif est distinct de la focalisation contrastive (absente en naxi)

L'ajout d'un ton H intensif est un processus intonatif distinct de la focalisation contrastive (parfois appelée *accent contrastif*, en anglais *contrastive stress*). La focalisation contrastive telle qu'elle apparaît typiquement en anglais (et, dans le domaine asiatique, en chinois) n'existe pas en naxi, non plus que dans de nombreuses langues, dans le domaine tibéto-birman et ailleurs (Trubetzkoy 1939 [1976]:242, qui nomme le phénomène « accent de phrase », le dit d'ailleurs présent « dans beaucoup de langues », non dans toutes). Cela est établi par le test suivant, qui consiste à tenter l'ajout d'un ton H d'intensification sur diverses syllabes de la phrase « Il ne devrait pas agir ainsi, il ne devrait pas faire ça » (phrase entièrement constituée de syllabes au ton M) : aucun des énoncés (29-31) n'est bien formé (ce que signale l'astérisque en début d'énoncé).

(29) * tʰɯ̄ ʈʂʰɤ̄ bē bē ´ mɤ̄ ā.
3e sg. déictique de manière part. adverbialisante agir + H intensif NEG. verbe de modalité : devoir

Le sens qu'aurait eu (29) si l'emploi du ton H intensif était acceptable ici aurait été : « Il ne devrait pas *agir* ainsi : il peut penser ce qu'il veut, dire ce qu'il veut, mais pas passer à l'acte ».

(30) * tʰɯ̄ ´ ʈʂʰɤ̄ bē bē mɤ̄ ā.
3e sg. + H intensif déictique de manière part. adverbialisante agir NEG. verbe de modalité : devoir

Le sens qu'aurait eu (30) si l'emploi du ton H intensif était acceptable ici aurait été : « *Lui*, il ne devrait pas agir ainsi (même si par ailleurs, venant de quelqu'un d'autre, ce serait acceptable) ».

(31) * tʰɯ̄ ʈʂʰɤ̄ ´ bē bē mɤ̄ ā.
3e sg. déictique de manière + H intensif part. adverbialisante agir NEG. verbe de modalité : devoir

Le sens qu'aurait eu (31) si l'emploi du ton H intensif était acceptable ici aurait été : « Il ne devrait pas agir *de cette façon-là ;* il aurait pu parvenir à ses fins autrement ».

Afin de compléter la caractérisation du ton H intensif, il peut être utile de préciser qu'il ne s'agit pas d'un marqueur de focus : si le *focus* est défini (dans l'esprit de Paris 1999:204, Xu Yi 1999, Delais-Roussarie, Rialland *et al.* 2002, Mazaudon 2003) comme la partie de l'énoncé sur laquelle porte l'assertion (ou l'interrogation, selon les cas), la différence entre le focus et le processus optionnel d'intensification par ajout d'un ton H ressort clairement : le statut focal n'est pas optionnel. L'un et l'autre ont néanmoins partie liée, dans la mesure où le ton H intensif ne s'associe qu'aux éléments qui ont statut focal, comme le montre sa distribution complémentaire avec le conditionnel/topicalisateur /sé/ : la phrase (32) ci-dessus peut être comparée avec la phrase (30).

(32) tʰɯ̄ ´sé ʈʂʰɤ̄ bē bē mɤ̄ ā.
3e sg. COND./TOP. déictique de manière part. adverbialisante agir NEG. verbe de modalité : devoir

« Lui, alors, il ne devrait pas faire ça (mais d'autres en revanche pourraient, le cas échéant) ».

(Note : le ton H qui suit /tʰɯ̄/ appartient au morphème de conditionnel/topicalisateur /´ sé/, issu d'un ancien disyllabe : /lɯ́ sé/, dont la première syllabe a laissé pour trace un ton H flottant.)

2.4.3. L'ajout de ton H intensif est distinct de l'accent d'insistance

Il importe également de distinguer l'ajout de ton H intonatif du phénomène qui depuis Coustenoble et Armstrong 1937 est dénommé *accent d'insistance* (étudié ci-dessus, section 2.3.2). Le ton H intensif est compatible avec l'accent d'insistance, et apparaît souvent associé avec lui. Ainsi dans 33a-b :

(33a) dɯ̄ sỳ wà.
un chose verbe copule
« C'est pareil, c'est la même chose. »

(33b) dɯ̄ sỹ wà.
(mêmes mots ; avec ajout d'un ton H intensif)
« [Mais si,] c'est la même chose ! »

L'énoncé (33a) expose un fait ; l'énoncé (33b) apparaît dans un contexte de désaccord entre interlocuteurs. Il n'est donc pas surprenant que (33b) soit prononcé avec un effort articulatoire accru (ses corrélats typiques incluant un allongement consonantique, une modification de la qualité de voix, un degré accru d'ouverture de la mâchoire au cours des voyelles et des mouvements de la langue amplifiés, des variations amplifiées de F_0, et une pression sous-glottique accrue). Mais ces traits phonétiques sont superposés à la séquence tonale LH dont est porteuse la syllabe /sỳ́/, ils ne sont pas commandés par cette séquence. Le fait ressort des cas où ton H intensif et accent d'insistance sont tous deux présents mais portent sur des syllabes différentes. En voici un exemple, tiré d'un récit au sujet des « suicides d'amour » autrefois courants : le mariage était très tôt décidé par la famille (alors que les jeunes gens concernés étaient encore enfants), mais les adolescents étaient libres de nouer des relations amoureuses, de sorte que beaucoup d'entre eux s'éprenaient d'une personne qui ne leur était pas destinée ; au moment où ils arrivaient en âge de se marier, les amoureux décidaient souvent de se suicider ensemble. Le récit enregistré raconte la chanson que se chantaient les amoureux résolus à mourir ensemble : « Allons-nous-en là où on récolte sans semer, où tout pousse sans qu'il soit besoin de labourer, où on a les tigres pour montures, où il n'y a ni mouches ni moustiques... » Les propositions relatives du naxi précèdent leur « tête » *(head* ; l'*antécédent* de la grammaire occidentale traditionnelle) ; la longue série de propositions relatives (qui occupe plus d'une demi-minute sur l'enregistrement) précède donc le nom « lieu », /bɤ̀/, qui est suivi par la forme verbale /bɯ̄ là/ (« aller » suivi de la particule indiquant l'invitation à faire quelque chose ; sens de l'ensemble : « allons »). Ces trois dernières syllabes sont prononcées /bɤ́ bɯ̄ là/, avec ajout d'un ton H intensif sur /bɤ́/ (ce qui rappelle l'emploi d'un intensificateur en anglais dans l'expression *the very place*), et avec un surcroît considérable d'énergie articulatoire sur les deux dernières syllabes, /bɯ̄ là/. (Sur l'enregistrement « Yulong », dans le dossier du locuteur M3, le passage se trouve à 4 mn 47 sec.)

Cet exemple illustre la différence entre l'intensification catégorielle (« phonologique », sommes-nous tenté de dire) par ajout d'un ton H, d'une part, et d'autre part l'intensification non catégorielle dont l'accent d'insistance représente une forme extrême.

2.4.4. Perspective typologique au sujet du ton H intensif

L'exemple de langues comme le naxi (et le bai, langue voisine, décrite par Xu Lin et Zhao Yansun 1984) qui possèdent un mécanisme catégoriel d'ajout de ton H pour la mise en valeur de syllabes aux tons L et M confirme la corrélation entre fréquence fondamentale et saillance informationnelle qui existe dans un grand nombre de langues (au sujet du wolof, qui ne suit pas cette tendance générale, voir Rialland et Robert 2003 ; d'autres langues d'Afrique, tonales et non tonales, sont similaires au wolof de ce point de vue). Il y a toutes chances pour que l'ajout d'un ton L à des fins d'intensification soit beaucoup plus rare dans les langues.

2.5. Bilan des expériences sur la composante pragmatique de l'intonation en naxi

Les réalisations *intonativement faibles* d'une syllabe ne possèdent pas de valeurs caractéristiques (pour F_0 pas plus que pour le quotient ouvert, le rapport consonne/voyelle ou l'intensité) : elles présentent une grande variabilité. Ces observations rappellent les propositions de Shih, Kochanski *et al.* 2001 pour le chinois mandarin : dans leur système de synthèse du mandarin, chaque syllabe se voit affecter un coefficient de force ; plus une syllabe est faible, plus elle est soumise à l'influence des syllabes environnantes : un ton bas faible après ton haut pourra ainsi présenter une fréquence fondamentale élevée.

> Effort is ultimately measured in physical units, while the communication error probability is dimensionless, so a scale factor is needed to make the two compatible for addition. This scale factor varies from syllable to syllable, and we identify it with the linguistic strength, or importance of each syllable. If a syllable's strength is large, the Stem-ML optimal pitch contour will closely approximate the tone's template, and the communication error probability will be small. (...) On the other hand, if the syllable is unimportant and its strength is small, the produced pitch will be controlled by other factors: neighboring syllables and ease of production. (Kochanski et Shih Chilin 2001:6 : « Hierarchical structure and word strength prediction of Mandarin prosody »)

En revanche, s'agissant des syllabes *intonativement fortes*, dont nous dirons qu'elles reçoivent une *intensification intonative* (voir l'entrée correspondante du Glossaire), **la description en termes de réalisation plus ou moins complète d'un gabarit tonal, telle que la proposent Shih, Kochanski *et al.* 2001, ne suffit plus**. En effet, ces auteurs n'envisagent qu'un type de réalisation forte de la syllabe : une réalisation canonique de son ton lexical, réalisation fidèle du gabarit tonal défini pour chacun des tons lexicaux. Matthew Chen érige cette conception en modèle phonologique, ne concevant les phénomènes pragmatiques (« emphasis and contrast ») qu'en termes de « déplacements d'accent », qui ont pour effet la réalisation affaiblie de certains des tons (Chen 2000). Ce modèle, utile dans l'état actuel des systèmes de synthèse de la parole, ne correspond pas à la variété des réalisations observées en parole naturelle. Les limites imposées par les outils existants de synthèse de la parole ne doivent pas amener les linguistes à délaisser cette variété.

Pour finir, rappelons que les variations locales qui ont ici retenu l'essentiel de l'attention ne révèlent qu'une partie de l'intonation. Les variations intonatives des autres parties de l'énoncé jouent bien sûr un rôle : la pause (ou allongement de la syllabe qui précède), phénomène largement représenté dans le corpus ; l'aplanissage intonatif du post-focus, particulièrement fort sur les syllabes qui suivent immédiatement le focus, qui affecte également l'articulation des consonnes et des voyelles : pour prendre l'exemple de l'énoncé (17), que le même informateur a prononcé à cinq reprises, l'initiale /ʈʂ^{h}/ du déictique perd son occlusion dans deux des réalisations, et se trouve voisée de part en part dans une

troisième ; la syllabe /pú/ subit un voisement de l'initiale dans trois des réalisations, phénomènes qui préludent à son ellipse dans la réalisation transcrite comme (17') (voir section 1.1).

Les expériences réalisées dans ce chapitre autorisent à conclure qu'**en naxi comme dans d'autres langues d'Asie comme le vietnamien et le chinois mandarin standard, la présence de tons lexicaux va de pair avec une relative abondance de phénomènes intonatifs locaux**. Le naxi appartient au groupe des langues à marquage intonatif du focus (autrement dit, à *morphème intonatif de focalisation*), groupe qui paraît largement majoritaire parmi les langues du monde. Le wolof (Rialland et Robert, *op. cit.*) est une langue sans tons et sans morphème intonatif de focus ; le naxi est une langue à tons qui présente pourtant des phénomènes d'intonation pragmatique.

Ces observations légitiment l'expérience proposée au chapitre III, laquelle consiste à étudier les modifications de mots anglais, vietnamiens et naxi (placés en phrase-cadre) lorsque le contexte d'énonciation indiqué aux consultants varie. Cette expérience suppose que dans ces trois langues, l'énonciateur possède la latitude de signaler l'importance d'un mot par une modification intonative locale (un *morphème intonatif d'insistance*, si le terme paraît plus éclairant). Cela est bien établi pour l'anglais, et pour le vietnamien (voir par exemple Doan Thiên Thuât 1999). Dans le cas du naxi, en revanche, l'état fragmentaire de la documentation existante rendait nécessaire cette phase exploratoire préalable.

Table des matières du chapitre III (« Etude comparée des modifications locales induites par l'intonation pragmatique en naxi, vietnamien et anglais »)

Commentaires de méthode sur la tâche de production 107
Utilité d'une vérification visuelle des mesures 110
Une limite : l'absence de vérification perceptive 110

1. Données anglaises 111
1.1. Expérience-pilote 111
1.2. Mots choisis pour l'expérience principale 113
1.3. Attentes au sujet de la réalisation des mots-cible 115
1.4. Locuteurs et déroulement de l'enregistrement 115
1.5. Bilan des enregistrements 116
1.6. Représentation visuelle des données 116
1.7. Tableaux de résultats pour les locuteurs principaux (M3 à M7) 116
1.7.1. F_0 moyenne : la syllabe accentuée est plus fortement modifiée 116
1.7.2. Evolution de F_0 : l'allure de la courbe change dans la moitié des cas 118
1.7.3. Le quotient ouvert est plus bas en lecture insistante 119
1.7.4. Longueur de la rime : variété des stratégies 120
1.7.5. Peu de variation des formants, augmentation d'intensité acoustique globale 122
1.7.6. Bilan des mesures sur les données anglaises 127

2. Données naxi 128
2.1. Matériau linguistique 128
2.2. Réflexion sur les phrases-cadres 132
2.3. Bilan des enregistrements 133
2.4. Représentation visuelle des données 134
2.5. Tableaux de résultats 137
2.5.1. Les valeurs moyennes de F_0 augmentent en lecture insistante 137
2.5.2. Evolution de F_0 : variété des stratégies 137
2.5.3. Quotient ouvert : cohérence des stratégies individuelles 138
2.5.4. La longueur de la rime varie peu 139
2.5.5. L'intensité acoustique globale est plus élevée en lecture insistante, mais la différence est moindre que dans les données anglaises 140
2.5.6. Bilan des différences entre conditions de lecture dans les données naxi 141

3. Données vietnamiennes 143
Rappel de l'enjeu scientifique de la prise en compte du vietnamien : 143
3.1. Matériau linguistique 144
3.2. Réflexion sur les phrases-cadres 146
3.3. Bilan des enregistrements et résultats 148
Les caractéristiques de qualité de voix ressortent plus nettement encore en lecture insistante (constriction glottale du ton B2, phonation modale des tons D1 et D2) 160

4. Synthèse des résultats : comparaison entre les trois langues 161
4.1. Importance des modifications locales 161
4.2. Relation entre fréquence fondamentale et intensité dans les trois langues étudiées 162
4.2.1. En anglais : fortes modifications de F_0 et de l'intensité ; en naxi et vietnamien : modification relativement plus élevée de l'intensité 162
4.2.2. Le rapport $\Delta I / \Delta F_0$ est plus bas lors d'un changement de ton lexical que lors d'un changement de lecture soignée à lecture insistante 163

Chapitre III. Etude comparée des modifications locales induites par l'intonation pragmatique en naxi, vietnamien et anglais

L'expérience rapportée dans ce chapitre consiste à comparer les mêmes mots sous deux conditions de lecture, appliquant un protocole similaire dans les trois langues afin de mettre en regard d'une langue à l'autre les effets locaux[1] de l'intensification intonative. Les observations se concentrent donc essentiellement sur l'étude de la réalisation phonétique de mots monosyllabiques ayant statut de *focus* (le *nucleus* de la tradition britannique), cela dans deux contextes d'énonciation qui diffèrent par le degré d'insistance qui porte sur le mot-cible.

Cette expérience aboutit au constat de différences entre les trois langues dans l'emploi des paramètres prosodiques. Ces différences au plan *intonatif* sont conformes à ce que laissait attendre la différence de structuration entre ces langues au plan de l'*accentuation lexicale*.

Les résultats sont exposés ci-dessous langue par langue : anglais (section 1), naxi (section 2), vietnamien (section 3), avant une comparaison synthétique des résultats (section 4). Cet exposé est précédé de trois commentaires qui s'appliquent au même titre aux données des trois langues.

Commentaires de méthode sur la tâche de production

La comparaison est riche d'enseignements dès le stade de l'élaboration des consignes aux locuteurs : **en naxi et vietnamien, un simple changement de particule finale modifie l'attitude énonciative véhiculée par une phrase, ce qui n'est pas le cas en anglais**. L'idée de comparer deux conditions de lecture provient d'une étude sur un point de tonologie vietnamienne : il s'agissait d'établir si certaines syllabes comportaient ou non une glottalisation ; les syllabes concernées ont été enregistrées dans deux phrases-cadres qui différaient par la particule finale, en vue d'établir que les faits essentiels (glottalisation régulièrement associée à un ton, absence régulière de glottalisation pour un autre) étaient stables d'un contexte intonatif à l'autre. Cette expérience (rapportée dans Michaud 2004a) a par la suite suggéré l'idée de comparer dans le détail les deux conditions de lecture, qui étaient les suivantes : **énonciation soignée (dans un contexte de type didactique) d'une part, énonciation insistante nuancée**

[1] Les modifications à l'échelle de l'énoncé (telles qu'élévation de la plage de F_0, changement de débit, modification de la ligne de déclinaison) ne sont pas abordées ; elles ne sont certes pas quantité négligeable, mais les hypothèses qui motivent la présente recherche ne rendent pas leur étude indispensable.

d'impatience adressée à un allocutaire (ou *co-énonciateur*, dans le cadre d'analyse énonciativiste) **qui faisait preuve de mauvaise volonté, d'autre part**. En naxi, de proches équivalents sont aisément trouvés, du fait que cette langue possède également des particules finales. **En anglais, il paraissait souhaitable de calquer ces mêmes contextes, afin que la comparaison soit aussi directe que possible. Or la même opposition ressortit, en anglais, au domaine paralinguistique**, tel qu'il est délimité par George Trager, David Crystal et John Lyons (Trager 1958 ; Crystal 1969 ; Lyons 1977), puisqu'elle ne se traduit pas par un changement des mots de l'énoncé, seulement du contexte d'énonciation. **Cette réflexion souligne la relativité de l'opposition entre linguistique et paralinguistique**, ce qui rejoint la critique de Robert Ladd (Ladd 1994) à l'endroit de Mary Beckman et Janet Pierrehumbert (Beckman et Pierrehumbert 1986), et la remarque d'Albert Di Cristo : « l'analyse linguistique doit intégrer des faits linguistiques et des faits para-linguistiques, contrairement à ce que cette dichotomie terminologique donnerait à penser » (Di Cristo 2004:103 ; voir également Rossi 1999:29).

Les deux conditions, présentées ci-dessus, sont désignées respectivement comme réalisation *soignée* (abrégée en S dans les tableaux et figures) **et réalisation *insistante*** (abrégée en I) ; **appellations choisies pour leur concision, qui n'ont pas valeur définitoire**. Le choix de ces conditions de lecture s'écarte des pratiques les plus courantes dans les laboratoires de phonétique, où l'échelle la plus fréquemment employée actuellement pour évaluer les réalisations d'un même item s'étend de la réalisation la plus *hypo-articulée* à la plus *hyper-articulée* (termes dont l'acception actuelle remonte à Lindblom 1990). Rapportées à ce continuum, nos conditions de lecture *soignée* et *insistante* se situent l'une et l'autre sur le versant hyper-articulé, ce qui peut laisser craindre qu'elles ne présentent un degré de similarité qui limite l'ouverture de champ de l'étude. Pour anticiper sur les résultats, ceux-ci montrent une différence significative entre les données recueillies dans les deux contextes d'énonciation, ce qui est de nature à rassurer au sujet de leur bonne différenciation. Il serait en outre problématique d'éliciter une prononciation hypo-articulée de mots présentés sous forme de liste : l'unique élément variant attire nécessairement l'attention ; une articulation affaiblie ne peut dès lors être imposée que par une consigne méta-linguistique, ce qui n'apparaît pas compatible avec la simulation d'un contexte de dialogue.

Un autre point par lequel cette expérience s'écarte quelque peu des procédures les plus courantes est que **le débit de parole n'a pas été contrôlé directement.** Les changements du débit de parole ont des effets importants sur l'articulation (voir par exemple Allen et Miller 1999 et références, et Smith 2002) ; il peut donc être indiqué de contrôler ce paramètre, par des consignes telles que « lire très lentement », « lentement », « vite », « le plus vite possible ». Notre choix tient, à nouveau, au fait que de telles consignes n'apparaissent pas

compatibles avec les indications de contexte[1]. Il est imaginable que le comportement des divers locuteurs aurait été plus proche si les consignes les avaient guidés avec plus de fermeté ; mais de telles consignes auraient accru la distance entre leur comportement en laboratoire et les emplois de la langue hors laboratoire. Ainsi, dans l'expérience de Pierrehumbert et Liberman 1984 qui consiste à faire varier le degré d'insistance *(emphasis)* sur une même phrase[2], et à relever les modifications de F_0 qui en résultent, les auteurs recourent à une indication chiffrée sur le degré d'insistance désiré :

> ...the pitch range instruction was varied in 10 steps, and six to eight repetitions of each pattern in each pitch range were recorded. In both of the experiments to be described, "degree of overall emphasis or excitement" was the term used in the subjects' instructions, and the kind of variation desired was illustrated by example. (Pierrehumbert et Liberman 1984:169)

Leur étude porte sur quatre locuteurs, dont les deux auteurs ; il est demandé aux deux autres sujets d'adopter le comportement dont les auteurs leur fournissent l'exemple (« For subjects other than the authors, the desired intonation patterns were demonstrated by example before the experiment, and the ability of the subjects to produce them naturally was checked » : Pierrehumbert et Liberman 1984:172). La fortune qu'a connu cet article tient pour une part à la clarté des résultats qui se dégagent : comparant les courbes de F_0 correspondant aux dix degrés d'insistance, le lecteur peut aussitôt constater que le point terminal de la courbe (valeur relativement basse, considérée par les auteurs comme la réalisation d'un ton L) reste presque constant, tandis que le point le plus élevé (modélisé comme un ton H) a une hauteur proportionnelle au degré d'insistance. Néanmoins, d'autres résultats sur l'anglais (dont ceux de Nolan 1995, qui a repris le principe de cette expérience ; ainsi que nos propres résultats, rapportés ci-dessous) ne vont pas dans le même sens ; la cause en est peut-être le protocole expérimental choisi par ces auteurs. Notre expérience en tant que sujet d'expériences (de production et de perception) suggère que le fait de se trouver dans un studio ou même dans

---

[1] Par acquit de conscience expérimentale, le débit a été contrôlé lors d'une troisième lecture, après la lecture sous les deux conditions décrites ci-dessus : tous les locuteurs anglais ont enregistré les mêmes mots dans la même phrase-cadre avec pour indication de lire « le plus vite possible ». Il leur a été précisé que cette troisième tâche, utile pour des raisons techniques, était distincte des deux autres et ne comportait pas d'indication de contexte. Au bilan, il est apparu que le compromis entre vitesse de réalisation et clarté de prononciation était nettement différent d'un locuteur à l'autre : énonciation à la fois rapide et très énergique chez certains, débit accru et articulation moins nette *(hypoarticulation)* chez d'autres. Des observations (non quantifiées) suggèrent que la saillance du mot-cible varie également selon les locuteurs, certains veillant à lui conserver un traitement particulier. En définitive, la comparaison porte uniquement sur les deux tâches contextualisées.

[2] Les consultants lisent la réponse des paires question-réponse qui leur sont présentées (sans indications complémentaires de contexte), du type de (A) et (B) ci-dessous :

What about Manny? Who came with him? (A) Anna came with Manny.

What about Anna? Who did she come with? (B) Anna came with Manny.

une salle d'hôpital, joint à la contrainte des appareils de mesure (invasifs ou non), n'a pas en lui-même une influence sensible sur le comportement, passées les premières minutes, tandis que **le détail des consignes données au sujet est déterminant pour la fiabilité des données recueillies**. Aucune expérience ne peut échapper entièrement au soupçon d'introduire des artefacts ; le degré d'artifice n'est pas indifférent pour autant.

Utilité d'une vérification visuelle des mesures

Le mode de traitement des données retenu ici comporte une vérification visuelle des résultats pour chacune des syllabes. Ce choix a pour prix de longues heures en tête-à-tête avec le détail des données. L'investissement de temps demeure toutefois raisonnable une fois rapporté au temps dévolu au projet de recherche, temps qui se compte en années de l'origine du projet jusqu'à la valorisation des résultats. Ce sentiment nous paraît rejoindre celui du comparatiste évaluant les mérites des procédures automatiques qui pourraient l'aider dans la reconstruction des langues (voir Mazaudon et Lowe 1991, Lowe et Mazaudon 1994 « The Reconstruction Engine: A Computer Implementation of the Comparative Method ») : en théorie, celles-ci permettent certes la prise en compte d'un nombre considérable de données, avec en outre l'avantage d'une application stricte des principes d'analyse définis au départ ; mais en pratique, cet outil (dans la version présentée par Lowe et Mazaudon 1994) impose à l'utilisateur d'écarter un à un les candidats non plausibles au statut de cognat, tâche ingrate qui demande au comparatiste de troquer provisoirement son statut de *Sherlock Holmes de la parenté linguistique* contre celui de *contrôleur de la qualité*, contraint de filtrer des rapprochements calculés automatiquement et durement confronté aux limites des modèles sémantiques appliqués dans les procédures de calcul. La réflexion à partir de statistiques fondées sur une masse de mesures non vérifiées en détail nous paraît présenter un inconvénient du même ordre, obligeant l'enquêteur à évaluer à chaque étape le degré d'incertitude des résultats et le risque d'artefacts.

Une limite : l'absence de vérification perceptive

Notre impression est que tous les locuteurs ont réalisé une différence nette entre les deux contextes d'énonciation qui leur étaient indiqués. **Les données recueillies appelleraient un test de perception pour vérifier que cette différence est bien perçue par des auditeurs natifs.** Ces tests de perception n'ont pu être menés pour le naxi et le vietnamien, car la présence de particules grammaticales différentes dans l'une et l'autre condition biaiserait les résultats obtenus par l'écoute de l'énoncé entier ; l'écoute de mots coupés de leur contexte phonétique immédiat serait également problématique. Pour l'anglais, le test n'a pas été mené pour des raisons pratiques.

1. Données anglaises

L'abondance des études sur la prosodie de l'anglais peut donner l'impression que cette langue peut désormais servir de pierre de touche dans les travaux de comparaison entre langues[1]. La prise de nouvelles données pourrait paraître superflue du fait de la documentation existante sur l'intonation des différents dialectes de l'anglais : par exemple, pour l'anglais parlé dans les îles britanniques, le corpus IViE d'Esther Grabe et Francis Nolan, et de nombreuses bases de données de *B.B.C. English*, dont le corpus MARSEC (Auran, Bouzon *et al.* 2004). L'impressionnant volume des études ne signifie pourtant pas qu'un accord ait été atteint sur le fond. Les résultats qui font actuellement autorité reposent rarement sur une base documentaire large. Lors de la recherche de résultats (ou de bases de données) qui répondent à un critère précis au plan expérimental, tel que la mesure du quotient ouvert, la masse des données ne livre souvent plus que des informations insuffisantes au propos d'une comparaison détaillée. **L'expérience présentée ici, à laquelle ont participé trois locuteurs en France (pour une série d'expériences-pilote) puis quatre locuteurs à Cambridge, peut paraître modeste, mais ses résultats ont plus qu'une valeur anecdotique**.

1.1. Expérience-pilote

La présentation des expériences-pilotes précède celle de l'expérience principale, et fournit l'occasion d'exposer le cheminement qui y a conduit.

1.1.1. Mots retenus pour la première étape (étude-pilote)

Deux ensembles de mots ont été constitués : des disyllabes et des monosyllabes. Le naxi et le vietnamien sont des langues monosyllabiques[2] ; le plus simple aurait été de n'employer que des monosyllabes dans les enregistrements d'anglais destinés à la comparaison. Ce choix aurait néanmoins soulevé une difficulté : en anglais, les caractéristiques prosodiques sont définies à l'échelle du mot, non à l'échelle des syllabes prises individuellement. N'employer que des monosyllabes serait négliger la caractéristique importante de la prosodie anglaise qu'est l'alternance entre syllabes accentuées et non accentuées.

---

[1] Voir, entre autres bibliographies récentes, celle de thèse de Caroline Bouzon (Bouzon 2004:310-326). La proportion des études sur l'anglais est très élevée dans les bibliographies générales sur l'intonation, par exemple celle actuellement disponible sur le site de l'université de Stuttgart : http://www.ims.uni-stuttgart.de/phonetik/joerg/biblio/biblio.html.

[2] Ce terme traditionnel demeure employé car les mots polysyllabiques observés dans ces langues sont des recompositions à partir d'éléments monosyllabiques (fait bien établi, rappelé récemment par Abramson, Thongkum *et al.* 2004, ainsi que par Larry Hyman en 2005 au séminaire LACITO-Sorbonne Nouvelle). Une discussion originale sur le processus historique qui aurait conduit au monosyllabisme en chinois à partir d'états antérieurs *polysyllabiques*, puis *disyllabiques* et enfin *sesquisyllabiques* est proposée par Ferlus 1998a.

1.1.2. Phrases-cadres

Dans un premier temps, les phrases-cadres du vietnamien ont été calquées littéralement :

This is the word __ . *(lecture soignée)*
This is the word __ , can't you hear? *(lecture insistante)*

Le premier de ces énoncés place le mot-cible en position finale, ce qui n'est pas souhaitable du fait des phénomènes d'allongement final (voir notamment Beckman et Edwards 1990) et de marquage intonatif de l'intonation de phrase (typiquement, pour cet énoncé affirmatif, une descente finale), qui interagit avec le phénomène qu'il s'agit au contraire d'isoler[1]. Le second de ces énoncés pose un problème de structure intonative : à la différence des particules finales du vietnamien et du naxi, le *checking-tag*[2] anglais constitue une unité intonative séparée ; de plus, l'attitude véhiculée revêt une certaine brutalité. De fait, le premier locuteur de l'expérience-pilote a ressenti de la difficulté à maintenir au long de la liste de mots l'intonation qu'appelle la deuxième phrase-cadre, "This word is pronounced ___ , can't you hear?" La recherche d'un équivalent des particules finales de langues à tons d'Asie nous amenait à choisir une phrase-cadre qui véhicule un message explicite d'agacement ou de colère, alors que l'insistance pouvait plus simplement être élicitée sans changer la phrase-cadre. Suite à cette étude-pilote, il a été décidé de varier les indications de contexte, non la phrase-cadre.

La différence que ce choix introduit entre les consignes données aux locuteurs anglais, naxi et vietnamiens fait d'emblée peser un léger doute sur les résultats de la comparaison entre l'anglais, d'une part, et les *langues à particules* que sont le naxi et le vietnamien, d'autre part : les locuteurs anglais étaient moins fermement guidés lors de leur tâche de lecture, puisque les phrases qu'ils disaient ne comportaient pas les particules qui, en naxi et vietnamien, inscrivaient, dans chaque énoncé, un indice de l'attitude énonciative à adopter. Il est donc imaginable que cela soit cause de la variabilité des résultats obtenus pour l'anglais. (Ce constat illustre les obstacles à la comparaison entre langues qui sont séparées par de profondes différences typologiques.)

Un adverbe *here* a été placé après le mot-cible, concession aux exigences du protocole expérimental : cet adverbe évite que le mot-cible ne soit en position finale ; il ne crée pas de frontière après le mot-cible puisqu'un adverbe en position finale est atonique[3]. Enfin, **les phrases ont été simplifiées en *Look, this is __ here***. L'emploi de l'impératif *look* s'est avéré

---

[1] La figure 13 (M2_mono) montre une convergence finale des valeurs de fréquence fondamentale dans les conditions S et I, effet probable de la position finale d'énoncé, qui limite la validité de la comparaison des deux conditions.

[2] Au sujet des notions de *tone units*, *tonicity*, *tones*, et du comportement des *tags*, voir en particulier Cruttenden 1986, ainsi que Deschamps, Fournier *et al.* 2004.

[3] Un des locuteurs (M7) a précisé après l'enregistrement qu'il n'aurait sans doute pas utilisé *here*, tout en signalant que sa présence ne l'avait pas pour autant gêné.

utile pour se rapprocher d'un style oral authentique. Il a été indiqué aux locuteurs que le choix de la phrase-cadre était motivé par des raisons techniques, sans leur fournir de précisions qui auraient pu divertir leur attention vers les sphères métalinguistiques. Il a été demandé de réaliser un cycle respiratoire (une inspiration et une expiration) entre chacun des 48 énoncés. Cette pause imposée visait à éviter l'*effet de liste* : elle devait conduire le locuteur à prendre le temps de se concentrer sur chaque énoncé, afin que chacun d'eux soit porté par une intention de communication, sans que le locuteur modifie sensiblement sa façon de lire au fil des 48 phrases. Enfin, il était indiqué au locuteur qu'il n'était pas nécessaire de faire un effort pour rompre la monotonie (introduire un élément oratoire, théâtral, comme il peut parfois être indiqué de le faire dans un cours de langue). Les consignes étaient les suivantes :

> ***Task 1:*** Context: You're teaching a foreign student who made a mistake when reading a word. (Class context.)
> Read each item inside the carrier sentence, making a long pause (breathing in and out once) between sentences.
> Look, this is __ here.
>
> ***Task 2:*** Context: A child who is learning to read has asked you how to pronounce this word time and again; (s)he asks you yet another time; you answer, less patiently:
> Look, this is __ here!
> Remember to make a long pause (breathing in and out once) between sentences.

Les enregistrements-pilotes ont suggéré que les deux conditions se distinguaient nettement par des différences locales de hauteur et de qualité de voix, plus manifestes sur la première syllabe (syllabe accentuée) que sur la seconde. Ces premiers résultats ont encouragé à passer au stade de l'expérience principale, avec un nombre accru de locuteurs.

1.2. Mots choisis pour l'expérience principale

Deux ensembles de 24 mots ont finalement été choisis :

- vingt-quatre monosyllabes de la forme CVt, où C est une consonne initiale occlusive (/p^h/, /p/, /t^h/, /t/, /k^h/, /k/, traditionnellement notées comme une série voisée et une série non voisée : /p/ /b/ /t/ /d/ /k/ /g/) ou une nasale (/m/, /n/) ; V une voyelle longue (/ɑ:/, /i:/, /u:/) ; et t la consonne finale /t/. Certains de ces mots sont répétés.
- vingt-quatre disyllabes ayant la même composition que les monosyllabes avec ajout d'un /ə/ : CVtə.

La liste complète des mots, mise en ordre aléatoire, est présentée dans le tableau 3.1.

Monosyllabes				Disyllabes			
Pete	coot	moot	beat	Peter	cooter	mooter	beater
poop	cart	teat	boot	Pooter	Carter	teeter	booter
mart	deep	toot	Garp	marter	deeper	tooter	garter
meet	Bart	tart	moot	meter	barter	Tartar	mooter
boot	geek	boot	mart	booter	geeker	booter	marter
part	dart	teat	neat	Parker	darter	teeter	neater

Tableau 3.1. Liste des mots enregistrés par les locuteurs anglais M3 à M7.

L'emploi de la consonne finale /t/ plutôt que /p/ ou /k/ tient à sa plus grande fréquence en position finale de syllabe, permettant de construire 20 des 24 paires monosyllabes/disyllabes recherchées (les autres paires recourant à /p/ ou /k/). En anglais, la consonne /t/ en position finale ou intervocalique est sujette à glottalisation, à un degré variable selon les locuteurs et les contextes (voir Przedlacka 2000 et références citées), ce qui a une incidence directe sur le mode de vibration des plis vocaux *(qualité de voix)*, et une incidence indirecte sur F_0, l'intensité et d'autres paramètres. La comparaison des mesures de quotient ouvert dans les deux conditions de lecture revient donc à enrichir l'étude d'une prise en compte de *la variation du degré de glottalisation* comme paramètre intonatif.

Le choix des voyelles longues les plus distantes à l'intérieur de l'espace articulatoire et acoustique offre le double avantage de permettre une étude de l'effet de la voyelle sur les paramètres étudiés (hauteur, qualité de voix), et inversement, de l'effet des conditions de lecture sur l'articulation vocalique (reflétée par la fréquence des formants). Le choix des consonnes vise également à permettre une estimation des caractéristiques co-intrinsèques de ces sons : l'électroglottographie offre des informations précises sur la mise en place du voisement après les occlusives. (L'existence d'une influence des consonnes sur la mélodie des voyelles est connue au moins depuis Henri Maspero [Maspero 1912] ; voir également House et Fairbanks 1953, Lehiste et Peterson 1961, Hombert 1978, et les travaux montrant que la perception de la consonne peut être changée par une modification de la mélodie de la voyelle (Fujimura 1971, Haggard, Ambler *et al.* 1970, Haggard, Summerfield *et al.* 1981.) L'effet des phonèmes sur les courbes de fréquence fondamentale et quotient ouvert des syllabes anglaises est étudié en détail dans l'Annexe 2, sections 1.3 et 1.4.

La liste comporte des mots de nature grammaticale variée : noms propres, noms communs, verbes et adjectifs. Leur fréquence est également très variable. Tous figurent dans le *Shorter Oxford English Dictionary*, mais deux d'entre eux sont rares : *moot* (« A piece of hard wood hooped with iron at each end, used in block-making. Also, a ring-gauge for shaping treenails cylindrically to the desired size ») et *coot* (« An aquatic bird, *Fulica atra*, belonging to the rail family, with black and dark grey plumage and the upper mandible extended backwards to form a white plate on the forehead »). Cela introduit une variable supplémentaire : la fréquence d'emploi d'un mot a une incidence sur sa prononciation, influence évidente dans les cas où elle aboutit à un changement phonémique catégoriel (combien parmi les locuteurs

de l'anglais savent que *Mrs*, /'mɪsɪz, -s/, provient d'un ancien *Mistress* ?), et qui exerce une action continue sur la langue. Le fait invite à la prudence dans la comparaison entre mots, mais n'a pas d'incidence directe sur la présente étude, qui se concentre sur la différence qui peut exister entre les deux conditions de lecture : **les mots sont les mêmes dans les deux conditions, de sorte que la symétrie est complète en dépit de l'hétérogénéité des mots sélectionnés.**

1.3. Attentes au sujet de la réalisation des mots-cible

Dans les termes de l'école anglaise d'études prosodiques (O'Connor et Arnold 1973:13, Cruttenden 1986:51), **la mélodie attendue sur les mots-cible est un *high-fall*.** L'expression correspondante dans la notation de Grabe, Post, Nolan et Farrar, qui recourent à une description en termes de tons et de frontières, est H* + L 0%, où H* désigne un ton haut attaché à la syllabe accentuée, L un ton bas, et 0% une frontière de groupe intonatif sans spécification tonale (« an intonation phrase boundary without a tonal specification ») (Grabe, Post *et al.* 2000:162).

1.4. Locuteurs et déroulement de l'enregistrement

Le locuteur M3 a participé à la dernière phase des expériences-pilote ; le protocole expérimental n'ayant pas été modifié par la suite, les données de ce locuteur sont incluses aux côtés de celles des quatre locuteurs enregistrés au Laboratoire de phonétique de l'Université de Cambridge[1], M4 à M7. Ces derniers ont été rétribués pour leur participation. **Tous quatre ont entre 20 et 21 ans ; la variété d'anglais qu'ils parlent est le « parler naturel des personnes cultivées »** (pour reprendre à Delattre une élégante formule aux allures d'oxymore) du sud de l'Angleterre[2]. (Les trois locuteurs—M1, M2, M3—qui ont participé aux expériences-pilotes parlent également l'anglais du sud de l'Angleterre.) L'électroglottographe employé est le même appareil que pour les enregistrements de vietnamien et de naxi (EG2-PC). Le matériel du laboratoire de Cambridge a été utilisé pour l'audio : micro Sennheiser MKH40, pré-ampli SYMETRIX SX202, enregistreur DAT SONY 60ES.

---

[1] Vifs remerciements à Francis Nolan et Jeff Potter (Laboratoire de phonétique de l'Université de Cambridge) pour leur concours dans la recherche de locuteurs et l'organisation des enregistrements.

[2] Ce parler est également désigné comme « Standard Southern British English », « Received Pronunciation », « B.B.C. English », dénominations dont chacune a ses nuances, son histoire et ses partisans ; Daniel Jones a d'abord parlé de « Public School Pronunciation » ; dans son dictionnaire de la prononciation anglaise (Jones 1917), il décrit « the Received Pronunciation » comme « the everyday speech of families of Southern English persons whose menfolk have been educated at the great public boarding schools ».

1.5. Bilan des enregistrements

La consigne de réaliser une longue pause entre énoncés successifs a dû être rappelée à plusieurs reprises au cours des enregistrements pour contrecarrer la tendance à une accélération du débit. La réalisation d'un cycle complet de respiration n'a pas été systématiquement imposée dans les cas où la pause paraissait suffisante pour que chaque énoncé forme une unité nettement distincte.

Au plan technique, le signal électroglottographique enregistré est de bonne qualité pour tous les locuteurs, de même que le signal audio.

1.6. Représentation visuelle des données

Les figures 3.1 à 3.12 présentent les courbes de F_0 et quotient ouvert pour toutes les syllabes du corpus anglais. Ces douze figures, toutes fondées sur le même mode de calcul, peuvent paraître répétitives ; il nous paraissait néanmoins plus parlant de présenter chacune de ces synthèses visuelles, plutôt que de retenir celles d'un seul locuteur ou de moyenner les mesures d'un locuteur à l'autre, modes de présentation qui tous deux auraient pour effet de noyer les différences entre ces courbes, dont chacune illustre une stratégie individuelle.

Les mesures sont réalisées sur la rime syllabique (choix justifié dans l'Annexe 2, section 1.3). Les figures visent à fournir un *type idéal*[1] de rime : plutôt que de représenter les courbes moyennes en n points également espacés dans le temps, ce qui s'éloigne de la division en cycles glottiques qui caractérise les données mesurées, il paraissait utile de les représenter comme si chacune était une rime unique, divisée en cycles glottiques. La procédure de création de ces simulations est décrite dans l'Annexe 2, section 4.4.

1.7. Tableaux de résultats pour les locuteurs principaux (M3 à M7)

1.7.1. F_0 moyenne : la syllabe accentuée est plus fortement modifiée

La syllabe accentuée (monosyllabe ou première syllabe de disyllabe) connaît un changement significatif de plage de F_0 chez quatre des cinq locuteurs. Ce changement varie dans la proportion de 1 à 3 selon le locuteur. La syllabe non accentuée est moins affectée, sauf chez le locuteur M7. L'écart-type est fréquemment plus élevé en lecture insistante, mais cela n'est pas systématique (dix cas sur quinze).

Chaque valeur du tableau 3.2 représente une moyenne sur 24 items. La différence entre les deux conditions fournit une estimation de la différence globale de F_0 entre contextes d'énonciation. Les différences ne sont pas fournies en Hz mais en pourcentages, du fait que

[1] L'expression est empruntée à la sociologie de Max Weber.

la perception humaine ne fonctionne pas de façon linéaire, ce que reflète l'échelle des tons musicaux, également fournie ici[1].

Le caractère significatif ou non des différences est évalué à l'aide d'un test-t (*two-tail t-test*), appliqué à chacune des colonnes de la matrice de résultats rééchantillonnés, chaque ligne correspondant à une rime syllabique[2]. Dans le cas de F_0, les ensembles à comparer sont de taille identique, et un test-t apparié est appliqué. En revanche, dans le cas du quotient ouvert, présenté plus bas, un test-t non apparié est employé dans les cas où certains points de données n'ont pas été calculés, pour les raisons exposées au chapitre I, section 1.2.3.1.

La mesure de la différence d'écart-type de F_0 d'une condition à l'autre vise à déterminer si la variabilité d'un item à l'autre est accrue dans une énonciation insistante.

Les résultats essentiels du tableau 3.2 sont repris sur le graphique 3.1 (page 492).

[1] Il convient néanmoins de préciser que le choix entre mesure en valeurs absolues (en Hz) et relatives (en tons musicaux) ne va pas tout à fait de soi ; il existe des tenants de la notation en Hz, qui défendent l'idée selon laquelle, dans la plage de fréquence fondamentale la plus courante en parole, une notation en valeurs absolues serait plus adéquate (Jacqueline Vaissière, séminaire de Doctorat). Pour cette raison, les valeurs en Hz ont été conservées sur toutes les figures.

[2] Les colonnes sont au nombre de 100 ; rappelons, comme il a été précisé plus haut (dans une note en section 2.2.2 du chapitre II), que le choix de rééchantillonner toutes les courbes en 100 points n'implique nullement qu'une valeur en un de ces 100 points ait une quelconque valeur par elle-même, puisque dans cette division en 100 points, chaque point correspond à moins d'un cycle glottique. Le choix d'employer 100 points, plutôt que 20, 10 ou même 3 tient au souhait de ne pas dégrader la précision des mesures en amont du résultat final, de même que dans les sciences physiques les résultats numériques ne sont arrondis qu'à l'étape du résultat et non à chaque étape du calcul, ce qui introduirait en fin de calcul une incertitude élevée.

locu-teur	type de syll.	F_0 moyen (Hz) et écart-type		différence : cond. I moins cond. S, en % (gauche) et val. approchante en tons musicaux (droite)		partie de la syllabe (exprimée en dizièmes de la durée syllabique) sur laquelle la différence est significative	diff. d'écart-type de F_0: cond. I comparée à cond. S, en %
		condition S	condition I				
M3	monosyll. accentué	162 (31)	169 (30)	4	0.4	0	-3
M4		148 (18)	179 (14)	21	1.7	9 premiers dizièmes	-23
M5		87 (7)	106 (15)	22	1.8	8 premiers dizièmes	+97
M6		92 (12)	105 (9)	14	1.2	9 premiers dizièmes	-26
M7		141 (16)	206 (32)	46	3.8	10	+103
M3	1e syll. de disyll. (accentuée)	182 (23)	182 (33)	aucune différence		0	+46
M4		166 (12)	196 (13)	18	1.5	10	+5
M5		109 (5)	140 (12)	28	2.3	10	+133
M6		91 (7)	115 (12)	26	2.1	10	+74
M7		171 (13)	254 (24)	48	4	10	+78
M3	2e syll. de disyll. (non accentuée)	114 (5)	111 (7)	-2	-0.2	4e au 10e	+22
M4		107 (4)	113 (5)	5.9	0.5	8 premiers dizièmes	+9
M5		77 (6)	84 (5)	10	0.8	10	-23
M6		71 (7)	77 (5)	8.4	0.7	10	-27
M7		108 (6)	170 (17)	58	4.8	10	+174

Tableau 3.2. Valeurs moyennes de F_0 dans les données anglaises.

1.7.2. Evolution de F_0 : l'allure de la courbe change dans la moitié des cas

Dans la moitié des cas, l'allure de la courbe de F_0 sur la syllabe-cible est significativement différente d'une condition à l'autre. Sur la syllabe non accentuée, la différence va toujours dans le sens d'une pente accrue en condition de lecture insistante. Ce phénomène n'est pas directement lié à la montée de plage de F_0 : en effet, chez le locuteur M3, les disyllabes ne sont pas réalisés avec une F_0 plus élevée en énonciation insistante, mais syllabe accentuée et syllabe inaccentuée présentent toutes deux une pente plus forte. Il paraît possible d'interpréter cela comme le signe d'une stratégie visant à renforcer le contraste entre les deux syllabes du disyllabe. Le locuteur M7, chez qui la pente de F_0 de la syllabe inaccentuée n'est pas significativement différente d'une condition à l'autre, recourt à une différence de pente de F_0 sur la première syllabe, descendante en lecture soignée, montante en lecture insistante. Les locuteurs usent de la pente de F_0 selon des stratégies opposées : descente accrue en condition de lecture insistante sur la syllabe accentuée initiale de disyllabe chez M3, montée chez le locuteur M7 en lecture insistante sur cette même catégorie de syllabes, s'opposant à une descente en lecture soignée.

A la différence des valeurs présentées dans le tableau 3.2, qui correspondent à des moyennes sur l'ensemble de la syllabe, celles du tableau 3.3 concernent l'évolution de F_0 au cours de la

syllabe[1]. Rappelons que le terme de « glissando » est réservé à la perception qu'ont les auditeurs des variations de F_0, et n'est pas employé pour caractériser les données de production telles que celles-ci. (L'approximation de l'évolution de F_0 proposée ici est relativement grossière. Sur l'interaction, au plan perceptif, de la plage de F_0 et des mouvements de F_0, voir notamment Cohen et 't Hart 1967, 't Hart, Collier *et al.* 1990, et plus récemment Rietveld et Vermillion 2003.)

locu-teur	type de syllabe	évolution de F_0 du 2e à l'avant-dernier cycle glottique (en tons musicaux)		statistiques sur la différence d'évolution de F_0 entre S et I	différence d'évolution (dans les cas significatifs): condition I comparée à S
		condition S	condition I		
M3	monosyll. accentué	-3.6	-3.8	$t_{23} = 0.84$, $p = 0.40$	
M4		-3.3	-3.1	$t_{23} = 0.63$, $p = 0.53$	
M5		-4.2	-4.5	$t_{23} = 4.3$, $p = 0.0001$	descente plus forte
M6		-2.2	-2.4	$t_{23} = 1.66$, $p = 0.10$	
M7		-3.4	-2.5	$t_{23} = -0.54$, $p = 0.59$	
M3	1e syll. de disyll. (accentuée)	-1.6	-2.1	$t_{23} = 1.66$, $p = 0.10$	
M4		-1.1	-0.9	$t_{23} = -0.03$, $p = 0.97$	
M5		-1.12	-1.14	$t_{23} = 1.79$, $p = 0.08$	
M6		-0.44	-0.63	$t_{23} = 2.12$, $p = 0.039$	descente plus forte
M7		-0.3	+0.5	$t_{23} = -3.13$, $p = 0.003$	S descendant, I montant
M3	2e syll. de disyll. (non accentuée)	-1.41	-1.8	$t_{23} = -2.6$, $p = 0.052$	
M4		-1.4	-2.1	$t_{23} = 7.1$, $p = 6 \times 10^{-9}$	descente plus forte
M5		(=)	-0.9	$t_{23} = 4.73$, $p = 2 \times 10^{-5}$	descente plus forte
M6		(=)	(=)	$t_{23} = 1.53$, $p = 0.13$	
M7		-0.9	-1.9	$t_{23} = 7.3$, $p = 3 \times 10^{-9}$	descente plus forte

Tableau 3.3. Evolution de F_0 au cours de la rime dans les données anglaises. Chaque valeur représente une moyenne sur 24 items.

1.7.3. Le quotient ouvert est plus bas en lecture insistante

Le traitement de la syllabe accentuée est similaire chez tous les locuteurs : lorsque le quotient ouvert varie, c'est dans le sens d'une *valeur plus basse en lecture insistante*, signe d'une voix plus *pressée* (voir le tableau 3.4, et le graphique 3.2 [page 492], qui présente la variation de quotient ouvert d'une condition de lecture à l'autre pour les trois catégories de syllabes). Le fait que ce phénomène soit plus net sur les monosyllabes est peut-être partiellement dû à une glottalisation plus nette du /t/ en position finale de mot qu'en position intervocalique, caractéristique qui serait mise en valeur en lecture insistante.

[1] Au plan technique, des mesures plus complexes seraient imaginables, telles qu'une mesure entre le point le plus haut et le point le plus bas de chaque courbe, ou une comparaison entre la forme globale des courbes dans leur intégralité, au travers d'une approximation par une courbe mathématique *(curve fitting)*. Ces procédures n'ont pas paru s'imposer pour le propos de la présente recherche.

Chez M5 et M6, la différence entre conditions qui s'observe dans le cas des monosyllabes se retrouve sur la syllabe inaccentuée des disyllabes ; chez M7, une diminution de quotient ouvert entre conditions de lecture est présente sur toutes les syllabes accentuées, tandis que la syllabe inaccentuée a un quotient ouvert nettement plus élevé en lecture insistante. Le fait que la syllabe inaccentuée donne lieu à des stratégies aussi divergentes entre locuteurs peut être rapporté à deux facteurs antagonistes : appartenant à un mot-focus, elle est potentiellement forte au plan prosodique ; non accentuée et succédant à une syllabe accentuée, elle est potentiellement faible au plan prosodique. Les locuteurs jouent sur les deux syllabes en vue de réaliser l'intensification intonative du mot.

locu-teur	type de syllabe	Oq moyen (%), et écart-type, cond. S	Oq moyen (%), et écart-type, cond. I	diff. entre cond. I et S (%)	partie de la syllabe (sur 100 échantillons) sur laquelle la différence est significative	écart-type : cond. I moins S (%)	différence (dans les cas significatifs): condition I comparée à S
M3	monosyll. accentué	56 (12)	57 (11)	-1.3	aucun	-9	pas de différence
M4		42 (6)	39 (7)	-7	13-48	+26	(plus bas)
M5		44 (3)	42 (4)	-3	37-42	+28	(plus bas)
M6		45 (4)	42 (5)	-7	57-92	+16	(plus bas)
M7		64 (10)	56 (9)	-13	26-63, 90-100	-14	plus bas
M3	1e syll. de disyll. (accentuée)	59 (6)	58 (8)	-1.4	aucun	+43	
M4		50 (6)	49 (4)	-2.4	aucun	-31	
M5		52 (3)	52 (4)	=	93-100	+36	
M6		49 (5)	48 (5)	-2	aucun	+15	
M7		63 (6)	58 (5)	-7	30-100	-24	plus bas
M3	2e syll. de disyll. (non accentuée)	63 (10)	65 (15)	+4	aucun	+44	
M4		46 (5)	47 (6)	+2	3-23	+29	plus élevé
M5		49 (4)	45 (3)	-9	9-85	-25	plus bas
M6		68 (11)	48 (7)	-30	1-100	-39	plus bas
M7		59 (8)	72 (5)	+22	1-100	-38	plus élevé

Tableau 3.4. Valeurs de quotient ouvert dans les données anglaises. Chaque valeur représente une moyenne sur 24 items.

1.7.4. Longueur de la rime[1] : variété des stratégies

La différence moyenne de longueur sur l'ensemble des syllabes est négligeable ; une légère tendance à un allongement des syllabes accentuées et un abrègement des syllabes inaccentuées ressort néanmoins (voir tableau 3.5, et graphique 3.3 [page 492]). Les différences moyennes sont inférieures à 25 ms dans tous les cas ; rapportées à la fréquence de vibration moyenne des plis vocaux dans les données concernées, cela ne représente pas plus de trois périodes ; dans la plupart des cas, la différence moyenne constatée est comparable avec l'incertitude de la mesure elle-même : comme il a été expliqué au chapitre I, la mesure de longueur se fait de la première fermeture glottique à la dernière fermeture

[1] Rappelons que la longueur de la rime est définie comme la durée qui sépare la première fermeture glottique détectée sur le signal électroglottographique de la dernière fermeture.

glottique ; or **la précision de la mesure elle-même est *à une période près*, non à quatre ou cinq millisecondes près** (voir Lehiste 1970:10-17, et Allen 1978, « Vowel duration measurement: a reliability study »).

La différence de longueur est statistiquement significative dans dix cas ; au plan perceptif, il est imaginable que cette faible différence ne soit pas en elle-même perceptible. Le seuil de perceptibilité (en anglais JND, Just-Noticeable-Difference) des différences de longueur varie en fonction de la méthode employée pour son évaluation ; Ilse Lehiste rapporte les valeurs suivantes : « ... in the range of the durations of speech sounds—usually from 30 to about 300 msec—the just-noticeable-differences in duration are between 10 and 40 msec » (Lehiste 1970:13 ; voir également Huggins 1971b, 1971a, et Klatt et Cooper 1975). La question se complique néanmoins du fait que des augmentations de durée, d'intensité et de fréquence fondamentale (ainsi que d'autres paramètres) qui prises isolément ne seraient pas perceptibles puissent être perçues lorsqu'elles ont lieu conjointement. Pour cette raison, le caractère significatif des variations observées au plan de la production est rapporté malgré les incertitudes qui pèsent sur leur pertinence perceptive en tant que paramètre isolé.

Le tableau 3.5 montre à nouveau une variété de comportements d'un locuteur à l'autre. Chez M3, les monosyllabes ne sont pas significativement affectés, les disyllabes sont allongés sur la première syllabe et raccourcis sur la seconde ; chez M4, les monosyllabes sont plus brefs en lecture insistante, et les disyllabes plus brefs par leur seconde syllabe ; chez M6, les syllabes accentuées sont allongées en lecture insistante, tandis que les syllabes non accentuées ne sont pas affectées de façon significative. Dans l'ensemble, la syllabe non accentuée a tendance à être plus courte en lecture insistante ; cette différence n'est pas corrélée avec celle de quotient ouvert ni de F_0.

locu-teur	type de syllabe	longueur moyenne (ms) et écart-type		différence: cond. I moins S, en ms	différence: cond. I moins S, en %	résultats du test-t sur la longueur (valeur de t_{23} et p)	différence (dans les cas significatifs): condition I comparée à S
		cond. S	cond. I				
M3	mono-syllabe accentué	164 (49)	186 (51)	+22	13	-1.48; 0.1458	pas de différence significative
M4		145 (47)	131 (41)	-14	-10	(valeur de t non relevée) $p = 9 \times 10^{-4}$	plus court
M5		166 (40)	166 (49)	(=)	(=)	(valeur de t non relevée) $p = 0.95$	pas de différence significative
M6		130 (31)	154 (37)	+24	+18	(valeur de t non relevée) $p = 8 \times 10^{-4}$	plus long
M7		158 (48)	171 (47)	+13	+7	-2.6; 0.014	plus long
moyenne		**153**	**162**	**+9**	**+6**		

M3	1e syll. de disyll. (accen-tuée)	148 (41)	162 (41)	+ 14	+ 10	-4.2; 3 x 10^{-4}	plus long
M4		107 (33)	109 (29)	(=)	(=)	-0.90; 0.38	pas de différence significative
M5		112 (30)	122 (30)	+ 10	+ 9	5.32; 2 x 10^{-5}	plus long
M6		113 (33)	129 (28)	+ 16	+ 14	-6.3; 1.8577 x 10^{-6}	plus long
M7		124 (3)	127 (3)	(=)	(=)	-1; 0.30	pas de différence significative
moyenne		**121**	**130**	**+9**	**+7**		
M3	2e syll. de disyll. (non accen-tuée)	154 (18)	136 (22)	-18	-12	3.0; 0.0058	plus court
M4		132 (31)	109 (26)	-23	-18	9; 5 x 10^{-9}	plus court
M5		118 (23)	107 (16)	-11	-9	3.4; 0.0023	plus court
M6		77 (21)	71 (15)	-6	-7	1.15; 0.26	pas de différence significative
M7		96 (20)	90 (22)	-6	-10	3.9; 7.75 x 10^{-4}	plus court
moyenne		**115**	**103**	**-12**			
moyenne générale		**130**	**131**	**(=)**			

Tableau 3.5. Longueurs syllabiques dans les données anglaises. Chaque valeur représente une moyenne sur 24 items.

1.7.5. Peu de variation des formants, augmentation d'intensité acoustique globale

La fréquence des formants varie peu d'une condition de lecture à l'autre, tandis que l'intensité acoustique est nettement plus élevée en lecture insistante.

Les résultats de formants et d'intensité acoustique globale sont présentés dans le même tableau du fait de la relation qui les lie, la configuration des formants influençant l'amplitude acoustique globale du signal (au sujet de ce phénomène d'*intensité intrinsèque*, voir en particulier Rossi 1971). Les figures 3.13 à 3.27 présentent l'évolution des trois premiers formants en fonction du temps pour chacune des voyelles (3.13 à 3.17 : voyelle /ɑ:/, 3.18 à 3.22 : voyelle /i:/, 3.23 à 3.27 : voyelle /u:/) et les locuteurs M3 à M7. La méthode d'estimation des fréquences formantiques est présentée au chapitre I, section 2.1. Les résultats retenus pour l'analyse statistique sont les valeurs données par le logiciel PRAAT après vérification manuelle (figures 3.13b, 3.14b, etc. ; les résultats bruts sont également présentés, sur les figures 3.13a, 3.14a, etc.). Les courbes d'intensité (locuteurs M4 à M7) sont présentées sur les figures 3.23 à 3.34 ; l'axe des ordonnées varie selon le locuteur, du fait que l'intensité acoustique globale n'est pas mesurée en décibels absolus ; en revanche, l'étendue de l'intervalle représenté est constant (30 dB).

La proportion de valeurs formantiques supprimées est indiquée dans le tableau 3.6 ; ces suppressions sont dues à des raisons techniques dans certains cas (pôles étiquetés à tort comme formants), aux caractéristiques acoustiques du signal dans d'autres (disparition de formants ou formants supplémentaires) : des précisions sont fournies au chapitre I, section 2.1.2.

		M3	M4	M5	M6	M7
monosyllabes	condition S	23%	13.0%	6.3%	3.1%	3.6%
	condition I	20%	8.3%	6.2%	0%	11.4%
disyllabes	condition S	10.8%	2.0%	1.1%	1.9%	1.6%
	condition I	14.5%	7.0%	0.3%	1.3%	12%

Tableau 3.6. Proportion de valeurs formantiques supprimées lors des mesures semi-automatiques sur les données anglaises

Les valeurs calculées pour chaque syllabe sont rééchantillonnées de la même façon que les courbes de F_0 et de quotient ouvert ; ceci revient à normaliser les durées des courbes des figures 3.13 à 3.27. La significativité des différences entre les courbes correspondant à la condition soignée (en bleu) et à la condition insistante (en rouge) a été évaluée par des tests-t appliqués en trois points de données : à 1/3, ½ et 2/3 de la syllabe, cela afin d'écarter les points de données proches du début et de la fin de la rime, où l'influence des consonnes est la plus forte.

Les figures correspondent aux données de monosyllabes ; les résultats des disyllabes sont comparables, comme le montrent les tableaux 3.7 et 3.8. Il n'a pas paru utile de doubler le nombre de figures en présentant séparément ces deux ensembles ; les moyenner n'a pas paru souhaitable.

Les courbes de formants montrent une grande constance, mise à part la perturbation en début et fin de voyelle attribuable à l'influence du contexte consonantique[1] : les voyelles /i:/, /ɑ:/ et /u:/ telles qu'elles sont réalisées par ces quatre locuteurs, dans les deux conditions de lecture étudiées, ne sont pas diphtonguées. Ce constat autorise à moyenner les valeurs sur l'ensemble de la syllabe.

[1] Ainsi, le second formant d'une voyelle suivant une consonne coronale, telle que /d/, aura tendance à partir d'une valeur proche de 1800 Hz pour rejoindre la valeur-cible de la voyelle ; chez les locuteurs qui réalisent /u:/ comme une voyelle d'arrière, au second formant bas, cette influence se traduit par une nette transition descendante du second formant en début de voyelle. (La désignation des transitions comme *montantes* ou *descendantes* se fait en considérant le vecteur-temps comme repère ; si en revanche on prenait pour repère la consonne, la transition de F_2 de /d/ à /u:/ serait décrite comme *montante* de la voyelle vers la consonne.)

locu-teur	type de syllabe	F_1 : valeurs en cond. S et I	signi-ficati-vité F_1	F_2 : valeurs en cond. S et I	signi-ficati-vité F_2	F_3 : valeurs en cond. S et I	signi-ficati-vité F_3	intensité : diff. en dB, et valeurs en conditions S et I	signifi-cativité intensité
M3	mono-	331 ; 338		2270 ; 2415		2900 ; 2975		+4.4 (65.8 ; 70.2)	1
M4	syllabe	301 ; 307		2280 ; 2280		3090 ; 2890	3	+3.1 (56.9 ; 60)	
M5		269 ; 266		2325 ; 2440	1, 2, 3	2800 ; 3020	1, 2, 3	+2.5 (55.0 ; 57.5)	
M6		284 ; 278	3	2200 ; 2300	1, 2, 3	3040 ; 3100	2, 3	+0.7 (63.8 ; 64.5)	
M7		294 ; 299		2300 ; 2350	2	2995 ; 2875		+4.2 (68.1 ; 72.3)	
M3	1e syll.	328 ; 346		2300 ; 2365		2850 ; 2890		+3.1 (68.6 ; 71.7)	
M4	de	301 ; 301		2280 ; 2265		2950 ; 2840		+2.3 (59.4 ; 61.7)	
M5	disyl-	271 ; 292	1	2393 ; 2390		2920 ; 2835	1, 2, 3	+4.7 (57.2 ; 61.9)	1, 2, 3
M6	labe	271 ; 300		2190 ; 2175		2975 ; 2940		+3.3 (65.9 ; 69.2)	1, 2, 3
M7		306 ; 288	1, 2, 3	2245 ; 2355	1, 2, 3	2920 ; 2815	2	+8.1 (70.4 ; 78.5)	1, 2, 3

Tableau 3.7. Résultats de mesures de fréquences formantiques et d'intensité sur la voyelle /i:/. Les points de mesure 1, 2 et 3 correspondent à 1/3, ½ et 2/3 de la durée de la rime. La colonne « significativité » indique ceux de ces points auxquels la différence indiquée dans la colonne qui précède est statistiquement significative.

La très faible variabilité du premier formant du /i:/ d'un item à l'autre et d'une condition de lecture à l'autre n'est pas due à un artefact de calcul (PRAAT et FOREST donnent des courbes extrêmement similaires). La quasi-constance de /i:/ est connue au plan articulatoire. Le phénomène peut s'expliquer par le caractère extrême de l'articulation de /i:/, qui contraint fortement sa réalisation, laissant peu de marge de variation[1].

Globalement, **F_2 est également stable. Les différences observées sont faibles (inférieures à 150 Hz dans tous les cas) ; dans les cas où elles sont significatives, elles vont dans le sens d'une augmentation de F_2** (antériorisation de la voyelle, soit une réalisation hyperarticulée de la cible formantique du F_2 du /i:/, qui est élevée). **Dans tous les cas, l'intensité acoustique est accrue en lecture insistante**.

[1] Voir Perkell et Nelson 1985, Kühnert 1996, et références citées. La variabilité intra-locuteur observée par Gunnar Fant, toutes voyelles confondues, est la suivante : de l'ordre de 25 Hz pour F_1, 50 Hz pour F_2, et 80 Hz pour F_3 (Fant 1974:41).

locuteur	type de syllabe	F_1 : valeurs en cond. S et I	significativité F_1	F_2 : valeurs en cond. S et I	significativité F_2	F_3 : valeurs en cond. S et I	significativité F_3	intensité : diff. en dB, et valeurs en conditions S et I	significativité intensité
M3	monosyllabe	511 ; 690	2	1115 ; 1160	3	2860 ; 2835		+7.3 (66.45 ; 73.75)	1, 2, 3
M4		651 ; 651		1126 ; 1143		2570 ; 2515		+2.3 (61.3 ; 63.6)	
M5		598 ; 608	1, 2, 3	1123 ; 1128		2690 ; 2730	1	+2.3 (55.9 ; 58.2)	1, 2
M6		566 ; 554		1110 ; 1140	1, 2, 3	2475 ; 2635	1, 2, 3	+0.4 (64.6 ; 65.0)	
M7		653 ; 675		1135 ; 1165		2755 ; 2880	1, 2	+4.1 (70.0 ; 74.1)	
M3	1e syll. de disyllabe	620 ; 660		1220 ; 1190	1, 2	2760 ; 2770		+5.5 (68.9 ; 74.4)	1, 2, 3
M4		645 ; 662		1170 ; 1180		2390 ; 2480		+0.5 (63.9 ; 64.4)	
M5		582 ; 578		1125 ; 1130		2700 ; 2750	1, 2, 3	+5.7 (57.8 ; 63.5)	1, 2, 3
M6		537 ; 563		1125 ; 1150	2	2400 ; 2590	2, 3	+2.9 (64.6 ; 67.5)	2, 3
M7		653 ; 651		1160 ; 1130	1, 2, 3	2900 ; 2950		+5.2 (71.8 ; 77.0)	2, 3

Tableau 3.8. Résultats de mesures de formants et d'intensité sur la voyelle /ɑ:/. Les points de mesure 1, 2 et 3 correspondent à 1/3, ½ et 2/3 de la durée de la rime. La colonne « significativité » indique ceux de ces points auxquels la différence indiquée dans la colonne qui précède est statistiquement significative.

Dans les cas où la variation du premier formant est significative, elle va dans le sens d'une augmentation : réalisation plus ouverte de /ɑ:/, ce qui correspond à une réalisation plus ouverte, réalisation hyperarticulée de la cible formantique du F_1 du /ɑ:/, qui est élevée. Aucune tendance commune aux divers locuteurs n'émerge pour le second formant. Dans l'ensemble, l'intensité acoustique est nettement plus élevée en lecture insistante.

locuteur	type de syllabe	F_1 : valeurs en cond. S et I	significativité F_1	F_2 : valeurs en cond. S et I	significativité F_2	F_3 : valeurs en cond. S et I	significativité F_3	intensité : diff. en dB, et valeurs en conditions S et I	significativité intensité
M3	monosyllabe	330 ; 336		1850 ; 1770	1	2390 ; 2410		+4 (68.6 ; 72.6)	2, 3
M4		310 ; 338	1, 2, 3	1015 ; 1100		2430 ; 2430		+2.1 (58.8 ; 60.9)	3
M5		291 ; 292	2, 3	1570 ; 1555		2236 ; 2216		+5.7 (55.3 ; 61.0)	1, 2, 3
M6		294 ; 298		1505 ; 1555	3	2160 ; 2290	1, 2, 3	+4.4 (65.7 ; 69.1)	3
M7		302 ; 280	2, 3	1510 ; 1325	2, 3	2200 ; 2235		+1.6 (72.0 ; 73.6)	
M3	1e syll. de disyllabe	336 ; 349		1820 ; 1800		2440 ; 2340	1, 2	+3.9 (69.7 ; 73.6)	1, 2, 3
M4		288 ; 314	1	900 ; 1125	1, 2, 3	2380 ; 2420		+1.6 (61.4 ; 63.0)	
M5		286 ; 295		1575 ; 1570		2235 ; 2275		+6 (58.0 ; 64.0)	1, 2, 3
M6		292 ; 299		1495 ; 1530		2170 ; 2260	1, 2, 3	+3.5 (67.7 ; 71.2)	1, 2, 3
M7		305 ; 272	1, 2, 3	1390 ; 1275	1, 2, 3	2200 ; 2310	1, 2, 3	+7.9 (72.4 ; 80.3)	1, 2, 3

Tableau 3.9. Résultats de mesures de formants et d'intensité sur la voyelle /u:/. Les points de mesure 1, 2 et 3 correspondent à 1/3, ½ et 2/3 de la durée de la rime. La colonne « significativité » indique ceux de ces points auxquels la différence indiquée dans la colonne qui précède est statistiquement significative.

Dans le cas de la voyelle /u:/ (tableau 3.9), l'intensité acoustique est dans l'ensemble nettement plus élevée en condition insistante. En ce qui concerne les formants, aucune tendance commune aux locuteurs n'émerge des données : F_1 est significativement plus élevé en condition insistante chez certains, significativement plus bas chez d'autres ; la même observation vaut pour F_2.

Les réalisations de /u:/ diffèrent d'un locuteur à l'autre. Rappelons que le second formant de la voyelle neutre, [ə], est théoriquement aux alentours de 1500 Hz ; chez M4, le second formant de /u:/ est nettement inférieur à cette valeur (proche de 1000 Hz) : /u:/ est réalisé comme une voyelle postérieure ; chez M5, M6 et M7, le second formant est proche de 1500, signe d'une voyelle proche de [ɯ] ; chez M3, le second formant est proche de 1800, signe d'une voyelle nettement antérieure, qui pour notre oreille francophone se rapproche d'un [y]. Les valeurs de F_3, qui chez tous les locuteurs contrastent nettement avec celles de /i:/, encouragent à penser qu'en anglais britannique contemporain, l'arrondissement est une caractéristique plus stable de la voyelle /u:/ que son caractère postérieur. Cela rappelle certains faits du russe (Martinet 1955:119). En diachronie, le passage de /u/ à /y/ est bien attesté dans des langues variées (Martinet 1955:52).

locuteur	F_1 : valeurs en cond. S et I	significativité F_1	F_2 : valeurs en cond. S et I	significativité F_2	F_3 : valeurs en cond. S et I	significativité F_3	intensité : diff. en dB, et valeurs en conditions S et I	significativité intensité
M3	538 ; 580	s	1613 ; 1600		2730 ; 2580	s	+4.2 (61.7 ; 65.9)	87
M4	588 ; 547	s	1480 ; 1620	s	2420 ; 2430		+0.4 (60.2 ; 60.6)	4
M5	524 ; 525		1615 ; 1670	s	2595 ; 2615	s	+3.8 (52.2 ; 56)	100
M6	357 ; 387		1585 ; 1725	s	2440 ; 2550	s	+5.1 (57.4 ; 62.5)	100
M7	495 ; 465	s	1650 ; 1780	s	2500 ; 2675	s	+8.3 (64.9 ; 73.2)	100

Tableau 3.10. Résultats de mesures de formants et d'intensité sur la voyelle réduite (/ə/) de la syllabe non accentuée des disyllabes. s = significatif ($p < 0.05$).

Du fait de la brièveté du /ə/, peu de points de données sont obtenus : moins de quatre dans environ 25% des cas. Le test statistique n'a donc été mené qu'en un seul point, à ½ de la syllabe.

La divergence des résultats pour /ə/ chez les divers locuteurs (premier formant plus élevé chez M3, plus bas chez M4 et M7) rejoint ce qui a été dit plus haut sur le statut de cette syllabe, faible par sa voyelle et son absence d'accent lexical, forte par la mise en valeur prosodique du mot auquel elle appartient. Au plan de l'intensité acoustique, en revanche, une augmentation nette en énonciation insistante est présente chez quatre des cinq locuteurs.

La moyenne des variations d'intensité entre conditions pour toutes les voyelles accentuées est présentée dans le tableau 3.11 en vue de la comparaison des valeurs d'intensité entre syllabes accentuées et syllabes non accentuées.

locuteur	Moyenne sur les rimes accentuées (monosyllabes, 1e syllabe de disyllabes) : condition I – condition S, en dB	Moyenne sur les rimes non accentuées (2e syllabe de disyllabe) : condition I – condition S, en dB
M3	+4.7	+4.2
M4	+2	+0.4
M5	+4.5	+3.8
M6	+2.5	+5.1
M7	+5.2	+8.3
Moyenne	**+3.8**	**+4.4**

Tableau 3.11. Bilan des différences d'intensité entre conditions de lecture dans les données anglaises

Chez les locuteurs M3, M4 et M5, l'augmentation d'intensité en condition de lecture insistante est plus nette sur la syllabe accentuée que sur la syllabe non accentuée ; chez les locuteurs M6 et M7, c'est l'inverse.

1.7.6. Bilan des mesures sur les données anglaises

En anglais, le passage de la *lecture soignée* à la *lecture insistante* se traduit de façon différente selon les locuteurs : augmentation d'intensité sous insistance, sans changement appréciable de la hauteur et de la qualité de voix ; augmentation d'intensité et hauteur accrue sur la syllabe accentuée, hauteur et qualité de voix inchangés sur la syllabe non accentuée ; ou augmentation d'intensité et hauteur accrue sur toutes les syllabes du mot-cible. Les variations de longueur, de pente de la courbe de F_0, et de position du pic de F_0 à l'intérieur de la syllabe varient également d'un locuteur à l'autre. La fréquence des formants change peu : augmentation significative des deux premiers formants du /u:/ chez un des locuteurs ; tendance à une augmentation du second formant pour toutes les voyelles, qui est significative dans ¼ des cas.

Dans le tableau 3.12, qui récapitule les différences entre conditions, un seuil a été choisi (empiriquement) concernant l'indication globale d'une différence de F_0, F_1, F_2 ou F_3 sur la syllabe : la différence est indiquée « s » si elle est significative sur plus des trois quarts de la durée de la rime. Pour F_0, ce choix se justifie par le fait que cette mesure vise à refléter la plage de fréquences dans laquelle la syllabe est réalisée, non les mouvements mélodiques.

Une case vide signale un facteur pour lequel aucune différence significative entre conditions n'a été relevée.

Lorsque le n-ème formant d'une voyelle est significativement affecté par le changement de condition de lecture, le symbole API de cette voyelle est reporté dans la colonne correspondante. Par exemple, le tableau signale que le F_2 du /u:/ change d'une condition de lecture à l'autre chez le locuteur M4.

Seules les syllabes accentuées sont décrites dans le tableau 3.12, car ce sont elles qui sont comparables avec les monosyllabes vietnamiens et naxi.

locuteur	type de syllabe	F_0 moyen (+ : cond. I > S ; - : I < S)	décroissance de F_0 (+ : cond. I > S ; - : I < S)	quotient ouvert: proportion de la rime sur laquelle la diff. est significative (+ : cond. I > S ; - : I < S)	longueur syllabique	F_1	F_2	F_3	intensité acoustique globale
M3	mono-syll.								/i:/ /ɑ:/ /u:/
M4		+		1/3 -	-	/u:/		/i:/	/i:/
M5		+				/ɑ:/	/i:/	/i:/	/ɑ:/ /u:/
M6		+		1/3 -	+		/i:/ /ɑ:/	/ɑ:/ /u:/	/u:/
M7		+	-	½ -	+	/u:/	/u:/		/i:/ /ɑ:/ /u:/
M3	1e syll. de disyll.		+	2/3 -	+				/i:/ /ɑ:/ /u:/
M4		+					/u:/		
M5		+			+			/ɑ:/	/i:/ /ɑ:/ /u:/
M6		+			+			/ɑ:/ /u:/	/i:/ /ɑ:/ /u:/
M7		+	-			/u:/	/ɑ:/ /u:/	/u:/	/i:/ /ɑ:/ /u:/
total		**8**	**3**	**1**	**6**	**1.3**	**2.3**	**2.6**	**7**

Tableau 3.12. Récapitulatif des différences statistiquement significatives entre conditions sur les monosyllabes anglais. La ligne total *indique la somme des cas de variation significative. Dans le calcul du nombre total de cas significatifs de variation des formants, le nombre de cas significatifs a été divisé par trois : ce calcul approximatif vise à rapporter le nombre de cas de variation significative au nombre total de syllabes.*

2. Données naxi

2.1. Matériau linguistique

En naxi, 120 mots ont été sélectionnés de façon à obtenir 40 triplets de mots de composition phonémique identique qui s'opposent par leur ton (H, M ou L), tels que /lá/ « frapper, heurter », /lā/ « tigre », /là/ « main ». Du fait de sa rareté lexicale, le quatrième ton lexical (ton montant LH) n'a pas été inclus. Tous les mots sont attestés dans le dictionnaire de Fang Guoyu et He Zhiwu 1995, et ont été vérifiés avant l'enregistrement avec chacun des cinq consultants[1]. Le choix des mots a été guidé par la nécessité de former des triplets ; la composition phonémique du corpus, exposée dans le tableau 3.13, représente un échantillon large mais non équilibré des combinaisons d'initiales et rimes permises par la langue naxi. (L'inventaire phonémique du naxi est présenté en détail dans l'Annexe 1.) Les consonnes initiales de ce corpus sont pour 30% des aspirées, pour 40% des non voisées, pour 30% des voisées.

[1] La principale différence dialectale entre ces cinq locuteurs est que l'initiale des items 106 à 108 est rétroflexe chez certains locuteurs (/ʂà/, /ʂā/, /ʂá/) et non chez d'autres, qui prononcent /sà/, /sā/, /sá/ ; il a paru légitime de considérer l'influence de cette différence consonantique sur la voyelle qui suit comme quantité négligeable.

	consonnes aspirées					consonnes non voisées						consonnes voisées				Ø	*tot.*
	p^h	t^h	k^h	ts^h	$ʈʂ^h$	p	k	ts	ʂ	s	h	g	l	m	n	Ø	*tot.*
i			x				x				x	x			x	x	18
y						x					x		x		x		12
ɯ		x		x			x	x			x		x				18
u					x									x			6
e				x									x				6
ɤ							x										3
o		x	x	x			x			x	x				x	x	24
ə˞					x				x		x						9
a									x								3
ɑ						x							x			x	9
wɑ					x												3
ɣ̩	x												x		x		9
tot.	3	6	6	9	9	6	12	3	6	3	15	3	15	3	12	9	120

Tableau 3.13. Composition phonémique du corpus naxi : les combinaisons initiale-rime employées sont marquées par des x ; « tot » désigne le nombre total de syllabes tonalisées (somme par lignes, et somme par colonnes)

La liste des mots est présentée dans le tableau 3.14. Les locuteurs lisaient la liste en ordre aléatoire.

Numéro assigné à la syllabe	con-son-ne	Notation API	Numéro dans le dictionnaire de Fang Guoyu et He Zhiwu 1995	Ordre dans la série lue
1	p	pỳ	1211	60
2		pȳ	1097	1
3		pý	434	120
4		pà	1173	16
5		pā	439	34
6		pá	901	59
7	p^h	p^hỳ	949	2
8		p^hȳ	452	98
9		p^hý	803	58
10	m	mù	524	61
11		mū	358	17
12		mú	514	35
13	h	hǐ	250	106
14		hī	446	36
15		hí	718	18
16		hỳ	773, 1183	99
17		hȳ	337	37
18		hý	559	3
19		hə̀	1185	108
20		hə̄	23	114
21		hə́	1088	38
22		hò	746	19
23		hō	927	96
24		hó	84	56
25		hù	14	39
26		hū	715	64
27		hú	125	104
28	l	lỳ	603	57
29		lȳ	503	20
30		lý	626	97
31		lù	1143	105
32		lū	79	40
33		lú	429	107
34		lỳ	753	4
35		lȳ	1071	95
36		lý	162	41
37		lè	3	21
38		lē	810	62
39		lé	939	115
40		là	722	42
41		lā	377	109
42		lá	967	63
43	t^h	t^hù	657, 768, 938	22
44		t^hū	558	5
45		t^hú	707	43
46		t^hò	1102	94
47		t^hō	191	47
48		t^hó	969	55
49	n	nò	276	44
50		nō	782	24
51		nó	1306	93
52		nỳ	257	103
53		nȳ	577	45
54		ný	713	6
55		ŋ̀ì	1195	54
56		ŋ̄ì	432	110
57		ŋ́ì	714	46
58		nỳ	60	48
59		nȳ	553	116
60		ný	582	65
61	g	gì	112	7
62		gī	805	25
63		gí	148, 917	53
64	k	kì	11	81
65		kī	1193	49
66		kí	1162	111
67		kɤ̀	891	52
68		kɤ̄	837	91
69		kɤ́	328	26
70		kò	790	67
71		kō	322	8
72		kó	857	117
73		kù	1101	50
74		kū	566	118
75		kú	683	66
76	k^h	k^hò	997	27
77		k^hō	343	92
78		k^hó	696, 1087	112
79		k^hi `	771	9
80		k^hi ¯	610	51
81		k^hi ´	599, 1116	29

82	ts	tsù[1]	1110	78
83		tsū	1122	90
84		tsú	641	10
85	ts^h	ts^hù	574, 1310	101
86		ts^hū	846	82
87		ts^hú	363	119
88		ts^hò	382	68
89		ts^hō	353, 637	79
90		ts^hó	960	89
91		ts^hè	1203	11
92		ts^hē	133	102
93		ts^hé	177	28
94	$ʈʂ^h$	$ʈʂ^h$ə̀	936	113
95		$ʈʂ^h$ə̄	642	69
96		$ʈʂ^h$ə́	732	83
97		$ʈʂ^h$ù	137	80
98		$ʈʂ^h$ū	915	70
99		$ʈʂ^h$ú	796	86
100		$ʈʂ^h$wà	931	30
101		$ʈʂ^h$wā	928	12
102		$ʈʂ^h$wá	1199	74
103	ʂ	ʂə̀	690	88
104		ʂə̄	1200	76
105		ʂə́	362, 1181, 903	23
106a		ʂà	*(1287)*	13a
107a		ʂā	*(749)*	31a
108a		ʂá	*(679)*	84a
106b	s	sà	*(1287)*	13b
107b		sā	*(749)*	31b
108b		sá	*(679)*	84b
109		sò	44	100
110		sō	97	73
111		só	640	85
112	Ø	ì	167	72
113		ī	385	14
114		î	568	32
115		à	686	71
116		ā	759	87
117		á	316	75
118		ò	317	77
119		ō	742	33
120		ó	905	15

Tableau 3.14. Liste des mots naxi enregistrés, arrangée par consonnes

[1] La réalisation phonétique de la voyelle des items 83 à 87 est [ʐ̩] : la rime /ɯ/ après une initiale fricative ou affriquée (/ts^h/, /ts/, /dz/, /ndz/, /s/, /z/) se réalise [ʐ̩].

2.2. Réflexion sur les phrases-cadres

2.2.1. Phrases-cadres et contexte d'élicitation pour les quatre premiers locuteurs (F2, M5, M7, M9)

Au cours d'un premier enregistrement réalisé en 2002, la phrase-cadre choisie était simplement la traduction en naxi de « Je dis __ » :

ŋɤ̀	__	ʂɤ́.
1e sg.	(mot cible)	dire

Le verbe étant en position finale, ces deux syllabes suffisent à encadrer la syllabe-cible. Les locuteurs, mis au courant de la nécessité de cette phrase-cadre, se sont pliés de bonne grâce à l'exercice. Afin d'obtenir une lecture insistante nuancée d'impatience, du même type que celle présentée ci-dessus au sujet des données anglaises (voir section 1.1 ci-dessus), il a été tenté d'ajouter une particule finale /mɤ̀/, qui se prête bien à l'expression de la contradiction :

ŋɤ̀ __ ʂɤ́ mɤ̀ !

que l'on peut gloser par « Mais enfin, je dis __ !/ Mais enfin, c'est __ que j'dis ! » Les consultants ont néanmoins jugé cette deuxième phrase-cadre difficile à énoncer dans le contexte de l'expérience.

Le contexte finalement imaginé pour la lecture *soignée* est celui d'un commentaire sur le travail d'un calligraphe en plein travail : plusieurs personnes regardent une émission télévisée sur la calligraphie ; un des spectateurs demande à son voisin quel est le caractère qui est en train d'être tracé ; il lui est répondu :

tʰɯ̄	ʈʂʰɯ̄	kʰɑ̀	__	ʈʂʰɯ̄	ndzɣ̀	pɤ̃	nɯ̀.
3e sg.	DEICT. moment *syllabe cible*			DEICT. caractère		écrire	TAM : aspect *en cours*

« Il est en train de tracer le caractère __ . »

L'idée d'imaginer le contexte d'une émission télévisée, plutôt que la situation (tout aussi plausible) d'une *séance présentielle* de calligraphie, a pour but d'éviter que les locuteurs ne se sentent tenus de chuchoter la réponse, réaction naturelle pour ne pas gêner le calligraphe au travail.

Dans la seconde condition (lecture insistante/impatiente), il est précisé que la personne qui pose la question est un enfant qui manque de politesse : il a déjà posé la question mais n'a pas prêté attention à la réponse ; la personne interrogée lui fournit à nouveau cette réponse, mais de façon à faire sentir à l'enfant que son comportement est irritant, et qu'il n'est plus question d'interrompre à nouveau. Une particule finale est ajoutée, pour refléter cette différence d'attitude : la particule affirmative /mɤ̀/, réduplicuée en /mɤ̋ mɤ̀/[1] :

[1] Au sujet de la réduplication en naxi, voir chapitre II, 1e partie, section 2.

$t^hū$ $ʈʂ^hū$ $k^hà$ __ $ʈʂ^hū$ ndzỳ pɤ̀ nù mɤ̀ˊ mɤ̀ !
(même énoncé que précédemment, particule finale mise à part)

2.2.2. Conditions spéciales pour le locuteur M8

Des consignes différentes ont été données au cinquième participant à l'expérience (le locuteur M8) : prêtre de la religion naxi (/tō mbà/), il pratique l'écriture pictographique traditionnelle, et gagne sa vie en produisant des objets destinés au marché touristique ; les phrases-cadres présentées ci-dessus le mettraient en position de simple spectateur du travail d'un autre ; il a paru indiqué, dans l'idée de lui faire pleinement jouer le jeu, de le placer dans le rôle du calligraphe, plus valorisant pour lui. La première phrase-cadre a été mise à la première personne :

ŋɤ̀	$ʈʂ^hū$	$k^hà$	__	$ʈʂ^hū$	ndzỳ	pɤ̀ nù.
1e sg.	DEICT.	moment *syllabe cible*		DEICT.	caractèreécrire	TAM : aspect *en cours*

« Je suis en train de tracer le caractère __ . »

Le contexte de cette énonciation *soignée* est alors le suivant : il est en train de travailler ; un visiteur l'observe, et lui demande quel caractère il est en train de tracer. La condition *insistante* se distingue de celle-ci de la même façon que précédemment : c'est un enfant malpoli qui pose la question ; l'artiste lui fournit à nouveau cette réponse, mais de telle sorte qu'il comprenne qu'il est en train de gêner, et qu'il n'est plus question d'interrompre à nouveau.

Une autre différence importante entre le locuteur M8 et les quatre autres est qu'il ne pratique pas l'alphabet naxi romanisé ; en revanche, il connaît les pictogrammes traditionnels. La liste des mots à enregistrer a donc été recopiée en pictogrammes par l'enquêteur, et revue signe par signe avec le consultant avant l'enregistrement ; lors de l'enregistrement, ce sont les mots transcrits par ces pictogrammes que le consultant a lus en les insérant l'un après l'autre en phrase-cadre.

2.3. Bilan des enregistrements

Cinq locuteurs ont participé à l'enregistrement : les locuteurs M5, M7, M8, M9 et F2, locuteurs natifs nés dans la région de Lijiang et séjournant dans cette ville, âgés de 30 à 50 ans[1]. L'enregistrement a eu lieu dans la ville de Lijiang, dans une pièce capitonnée par nos soins (à l'aide de matelas posés sur le sol et contre les murs, et de couvertures couvrant presque toutes les surfaces) afin d'atténuer les bruits extérieurs et de limiter les phénomènes d'écho. (Les fichiers sonores figurent dans un sous-dossier « mots sous deux conditions de lecture » du dossier des locuteurs concernés.) Dans l'ensemble, les résultats des enregistrements électroglottographiques montrent une **grande cohérence dans le**

[1] Au plan professionnel, M7 et M9 sont employés dans des musées (où ils effectuent des tâches de recherche et d'administration), F2 et M5 dans l'administration.

comportement de chacun des locuteurs, qui répètent des schémas intonatifs similaires sur des syllabes dont la composition segmentale varie. La différence entre les données produites sous les deux conditions de lecture (*lecture soignée* et *lecture insistante*) est statistiquement significative dans tous les cas, même si les paramètres qui varient significativement ne sont pas les mêmes d'un locuteur à l'autre.

2.4. Représentation visuelle des données

2.4.1. F_0 et quotient ouvert

Une même échelle a été employée pour toutes les figures[1]. Les figures 3.31 à 3.45 présentent les résultats locuteurs par locuteur, en suivant les mêmes conventions que pour les données anglaises et vietnamiennes. Comme il a été indiqué ci-dessus (section 1.6) au sujet des données anglaises, ces figures, toutes fondées sur le même mode de calcul, peuvent paraître répétitives ; il nous paraissait néanmoins plus parlant de présenter chacune de ces synthèses visuelles, plutôt que de retenir celles d'un seul locuteur ou de moyenner les mesures d'un locuteur à l'autre, modes de présentation qui tous deux auraient pour effet de noyer les différences entre ces courbes, dont chacune illustre une stratégie individuelle. Les mêmes conventions de représentation sont adoptées pour les trois langues étudiées, uniformité qui a pour effet de tripler le nombre de ces planches de figures quelque peu monotones, mais nous paraît utile pour faciliter la comparaison entre les données des trois langues.

2.4.2. Formants

Du fait du petit nombre des voyelles représentées dans les données (douze voyelles), la représentation de toutes les voyelles dans un espace F1-F2 et F2-F3 a paru plus adéquate que des figures séparées[2]. Les figures 3.46 à 3.55 montrent ces résultats.

L'absence de certains points de données tient au fait que les fréquences formantiques des voyelles concernées n'ont pu être estimées avec certitude. La proportion de valeurs formantiques supprimées est indiquée dans le tableau 3.15 ; comme il a été rappelé au sujet des données anglaises, ces suppressions sont dues à des raisons techniques dans certains cas (pôles étiquetés à tort comme formants), à des difficultés plus fondamentales dans d'autres (disparition de formants ou formants supplémentaires) : voir les rappels théoriques au chapitre I, section 2.1.2. Les difficultés d'estimation concernent en particulier la distinction

[1] Un léger compromis a été réalisé pour le locuteur M8 : la longueur syllabique étant très élevée, l'adoption de l'axe des abscisses correspondant réduirait la lisibilité des résultats des autres locuteurs, qui se trouveraient resserrés dans la partie gauche de la figure. Un axe des abscisses différent d'environ 10% dans son étendue a été employé pour les données de M8.

[2] Au sujet de la représentation et de l'analyse des systèmes vocaliques, voir notamment Schwartz, Boë *et al.* 1997b:236 ; Schwartz, Boë *et al.* 1997a:267-273, « The dispersion-focalization theory of vowel systems », et les propositions de Gunnar Fant (Fant 1960).

entre les deux premiers formants du /u/ et la détection du F3 (souvent d'amplitude très faible) chez le locuteur M7, le troisième formant du /i/ chez le locuteur M8.

	F2	M5	M7	M8	M9
condition S	3.4%	7.3%	6.8%	7.0%	4.6%
condition I	4.4%	7.1%	3.6%	4.9%	1.4%

Tableau 3.15. Proportions de valeurs formantiques (points de données) supprimées dans les données naxi

Du fait du petit nombre de données disponible pour chacune des voyelles, il n'a pas été tenté d'analyse statistique des fréquences formantiques. La représentation des deux premiers formants est réalisée sur une figure de forme carrée par commodité technique ; il est connu que les possibilités articulatoires appellent une représentation trapézoïde, voire triangulaire.

En naxi, le son [v] appartient au même paradigme que les voyelles : le phonème /v/ (dont la variation allophonique est décrite dans l'Annexe 1) apparaît exclusivement en position de noyau syllabique (notion qui en naxi coïncide avec celle de *rime* syllabique, la notion de consonne coïncidant avec celle d'*initiale* syllabique). Le son [z] correspond, soit à l'initiale fricative voisée /z/, soit à la voyelle /ɯ/ après une initiale fricative ou affriquée (/ts^h/, /ts/, /dz/, /ndz/, /s/, /z/). La notation classique pour ces sons consiste à souscrire un trait vertical : [ɣ̩], [ʐ̩], pour signaler que ces sons ont statut de noyau de syllabe. Sur les figures 3.46 à 3.55, ce diacritique a été omis par souci de clarté visuelle.

L'examen des courbes de formants syllabe par syllabe montre que les voyelles ne sont pas diphtonguées (à l'exception du /e/ qui a parfois tendance à se fermer en fin de syllabe, se rapprochant de [ej]), **ce qui nous a paru autoriser à moyenner les mesures de formants sur l'ensemble de la syllabe, et à représenter chaque voyelle comme un point dans l'espace vocalique**. Comme sur les autres figures, le bleu correspond à la condition de lecture soignée, le rouge à la condition insistante. Outre la répartition en voyelles, les items ont été distingués par le ton, en dessinant en grands caractères les mesures sur les syllabes au ton haut, en taille moyenne les syllabes au ton moyen, et en très petits caractères les syllabes au ton bas. Cela offre la possibilité d'évaluer visuellement l'influence du ton sur l'articulation de la voyelle.

Il nous semble que **les triangles vocaliques en deux dimensions (F_1-F_2), très courants dans les descriptions phonétiques, ne fournissent pas une image complète des systèmes vocaliques**. La prise en compte du troisième formant est utile pour faire ressortir clairement la distance qui sépare /i/ et /y/, par exemple, voyelles qui paraissent extrêmement proches lorsqu'elles sont présentées dans un espace F_1-F_2. La prise en compte du troisième formant a une importance particulière pour les langues, comme le français, dans lesquelles le trait d'arrondissement a valeur distinctive ; la représentation des voyelles rhotiques (en anglais *rhotic vowels* ou *roticized vowels*) présentes dans des langues comme l'anglais américain, le

pékinois et le naxi appelle également une prise en compte du F_3, puisqu'un F_3 très bas (au-dessous de 2 KHz) est la principale caractéristique de ces voyelles (par exemple le /ɚ/ du naxi).

Une solution proposée par Bladon et Fant consiste à calculer F_2' (F_2 prime), moyenne pondérée de F_2 et F_3 (Bladon et Fant 1978 ; solution également adoptée par Ao 1990). Plutôt que d'adopter cette solution technique, il nous a paru intéressant de représenter les trois premiers formants. La figure 3.56 montre une représentation tridimensionnelle des voyelles du locuteur naxi M3, qui fait ressortir la distance qui sépare /ɚ/ des autres voyelles, et, parmi ces dernières, la distance entre /i/ et /y/. Une représentation en mouvement dans un espace tridimensionnel F_1-F_2-F_3 paraît tout à fait parlante, car elle permet d'apprécier la configuration de l'espace vocalique, en variant les points de vue ; un script MATLAB a été réalisé à cet effet[1].

La prise en compte du troisième formant est moins nécessaire pour les voyelles postérieures. L'estimation précise de la fréquence de ce troisième formant est en outre difficile dans le cas de ces voyelles du fait de sa faible amplitude. Ainsi, les valeurs relativement dispersées du F_3 de /u/ sur les figures 3.46 à 3.55 reflètent la difficulté de la mesure de ce formant plutôt qu'une variabilité au plan articulatoire. Les valeurs de F3 de toutes les voyelles ont néanmoins été représentées, par souci d'exhaustivité.

2.4.2.1. Comparaison entre tons lexicaux

Aucune tendance générale concernant l'influence des tons sur la fréquence des formants n'émerge des données. Ce résultat est conforme à ce que laissait prévoir l'indépendance des voyelles et tons naxi au plan phonologique. Il infirme l'hypothèse d'une corrélation entre F_0 et F_1 formulée par A. Syrdal et H. Gopal (Syrdal et Gopal 1986) ; ce fait sera repris dans la discussion.

2.4.2.2. Comparaison entre conditions de lecture

L'examen des figures montre une **légère tendance générale à l'augmentation de F_1 et F_2 en condition de lecture insistante**, mais la variation est de faible ampleur. Le troisième formant est assez stable. Dans le détail, il existe des contre-exemples : chez le locuteur M8, la tendance serait plutôt inverse ; le /ɑ/ du locuteur M7 a un second formant légèrement plus bas en lecture insistante ; chez la locutrice F2, le /y/ a un premier formant plus bas dans cette même condition ; de façon générale, le premier formant des voyelles postérieures fermées varie peu d'une condition de lecture à l'autre.

---

[1] Le script que nous avons créé pour réaliser un triangle vocalique animé en trois dimensions est disponible sous forme d'exercice d'introduction au langage MATLAB, à l'adresse suivante : http://www.cavi.univ-paris3.fr/ilpga/ED/student/stam/page2b.htm

2.5. Tableaux de résultats

2.5.1. Les valeurs moyennes de F_0 augmentent en lecture insistante

locuteur	ton	F_0 moyen (Hz) et écart-type		différence : cond. I moins S, en % (gauche) et valeur approchante en tons musicaux (droite)		partie de la syllabe (sur les 100 échantillons également espacés) sur laquelle la différence est significative	diff. d'écart-type de F_0: cond. I comparé à S, en %
		condition S	condition I				
F2	H	240 (8.4)	274 (11.6)	+14	+1	1 à 99	+38
M5		169 (9.4)	179 (18.0)	+5.9	+1/2	8 à 89	+91
M7		247 (15.2)	277 (13.7)	+12	+1	2 à 100	(=)
M8		173 (8.8)	171 (6.8)	(=)	(=)	53 à 56, 58 à 67, 77 à 90, 92 à 98	(=)
M9		186 (14.0)	211 (12.9)	+13	+1	totalité (1 à 100)	-8
F2	M	207 (9.6)	225 (7.6)	+8.7	+3/4	1 à 99	(=)
M5		131 (6.2)	129 (8.5)	-1.5	(=)	89 à 100	(=)
M7		195 (14.3)	222 (9.8)	+13.8	+1	2 à 100	-46
M8		130 (7.6)	139 (6.9)	+6.9	+1/2	2 à 100	(=)
M9		128 (8.4)	159 (8.5)	+24	+2	2 à 100	(=)
F2	L	178 (13.9)	183 (8.2)	+2.8	+1/4	8 à 15, 17 à 21, 26 à 30, 59 à 64, 79 à 82	-59
M5		111 (4.8)	109 (5.1)	-1.8	(=)	7 à 20 : plus haut en cond. I ; 57 à 100: plus bas en cond. I	(=)
M7		147 (12.6)	163 (9.6)	+10.9	+1	2 à 97	-31
M8		105 (6.4)	109 (5.0)	+3.8	+1/3	3 à 82	(=)
M9		106 (5.6)	123 (5.1)	+16	+1 1/3	2 à 100	(=)

Tableau 3.16. Valeurs moyennes de F_0 des données naxi

La fréquence fondamentale moyenne augmente en lecture insistante chez tous les locuteurs, pour tous les tons. Aucun contre-exemple n'a été observé, seulement trois cas de stabilité d'une condition de lecture à l'autre (ton H du locuteur M8, ton M et ton L du locuteur M5). L'écart-type n'augmente pas de façon nette en lecture insistante.

2.5.2. Evolution de F_0 : variété des stratégies

Les variations de pente, à la différence des variations de plage moyenne de F_0, révèlent des stratégies variables selon les locuteurs, et selon les tons. La situation la plus commune est que la pente observée en lecture soignée devienne plus forte en lecture insistante, sauf chez le locuteur M9, chez qui la pente (descendante) du ton M et du ton L est moindre en lecture insistante. Cette variabilité tend à confirmer l'absence de spécification du contour des trois tons H, M et L du naxi.

L'évolution de F_0 est estimée de façon automatique en retenant deux valeurs pour chaque rime : le deuxième cycle glottique, et l'avant-dernier. Le choix d'écarter les deux valeurs extrêmes vise à limiter la variabilité due à l'effet de la consonne précédente et aux irrégularités parfois présentes en fin de voisement.

locuteur	ton	évolution de F_0 du 2e à l'avant-dernier cycle glottique (en tons musicaux)		statistiques sur la différence d'évolution de F_0 entre cond. S et I	différence d'évolution (dans les cas significatifs): condition I comparée à S
		condition S	condition I		
F2	H	+0.12	-0.15	s (t_{37} = 2.11, p = 0.038)	montée en cond. S, décroissance en cond. I
M5		+0.40	+0.56	(t_{37} = -1.06, p = 0.29)	
M7		-0.91	-0.93	(t_{43} = 0.45, p = 0.65)	
M8		-1.30	-1.61	s (t_{43} = 2.07, p = 0.041)	décroissance plus forte
M9		(=)	+0.47	s (t_{45} = -3.77, p = $3x10^{-4}$)	montée plus forte
F2	M	-0.32	-0.54	s (t_{42} = 3.66, p = $4x10^{-4}$)	décroissance plus forte
M5		-0.52	-1.06	s (t_{37} = 3.7, p = $3x10^{-4}$)	décroissance plus forte
M7		-2.12	-1.97	(t_{44} = 0.14, p = 0.88)	
M8		-1.95	-1.96	(t_{45} = 0.65, p = 0.52)	
M9		-0.66	(=)	s (t_{46} = -4.87, p = $4x10^{-6}$)	décroissance annulée
F2	L	-1.11	-0.95	(t_{44} = -0.56, p = 0.57)	
M5		-1.14	-2.16	s (t_{36} = 7.7, p = $5x10^{-11}$)	décroissance plus forte
M7		-2.29	-2.46	s (t_{39} = 2.13, p = 0.036)	décroissance plus forte
M8		-1.87	-2.26	(t_{42} = 1.79, p = 0.076)	
M9		-1.44	-0.80	s (t_{49} = -5.4, p = $5x10^{-7}$)	décroissance moins forte

Tableau 3.17. Evolution de F_0 dans les données naxi

2.5.3. Quotient ouvert : cohérence des stratégies individuelles

Chez certains locuteurs, le quotient ouvert est plus bas sous insistance, chez d'autres il est plus élevé, sans que cette variation puisse être rapportée à la plage de F_0 dans laquelle ces tons sont réalisés. Une certaine cohérence intra-locuteur s'observe dans l'emploi de la qualité de voix, signe d'une variété de stratégies individuelles. Ces observations sont en conformité avec l'absence de spécification phonologique de la qualité de voix des tons du naxi, qui laisse libre jeu aux stratégies individuelles[1].

[1] Le fait que les différences significatives de quotient ouvert s'observent plutôt en début de rime qu'en fin de rime (fait qui ressort plus clairement du tableau 3.21 que du tableau 3.18) nous paraît attribuable à un relâchement final qui affecterait les rimes d'une façon relativement indépendante de la condition de lecture, et se traduirait par des valeurs de quotient ouvert plus variables en fin de rime.

locu-teur	ton	quotient ouvert moyen (%)		différence entre cond. I et S (%)	partie de la syllabe (sur les 100 échantillons également espacés) sur laquelle la différence est significative	écart-type en cond. I comparé à cond. S (%)	différence (dans les cas significatifs): condition I comparée à S
		cond. S	cond. I				
F2	H	57.4	56.5	-1.6	22 à 46, 49 à 62	-18	
M5		55.4	58.7	+6.0	3 à 61, 90 à 100	+27	plus élevé
M7		59.5	56.8	-4.6	4 à 67, 87 à 94	+13	plus bas
M8		50.9	49.4	-1.0	87 à 99	-19	
M9		54.7	53.4	-2.4	75 à 93	-19	
F2	M	51.6	53.5	+3.7	0 à 20, 90 à 100	-23	
M5		59.2	60.3	+1	10 à 64 (plus élevé en cond. I), 74 à 94 (plus bas en cond. I)	+33	plus élevé au début, plus bas à la fin
M7		56.5	55.4	-2	7 à 27	-14	
M8		45.7	45.7	(=)	aucun	+3.9	
M9		53.8	51.6	-4	3 à 47	-30	plus bas au début
F2	L	49.0	50.0	+2	aucun	-15	
M5		61.8	65.7	+6	4 à 66	+70	plus élevé
M7		51.7	47.9	-8	6 à 93	-13	plus bas
M8		46.0	44.4	-3.6	9 à 12, 30 à 35, 38 à 69, 74-75, 86 à 98	+3	(plus élevé)
M9		54.7	50.8	-7.7	2 à 100	-44	plus bas

Tableau 3.18. Quotient ouvert des données naxi

2.5.4. La longueur de la rime varie peu

La variation moyenne de longueur est faible : 11 millisecondes, rapportées à la fréquence de vibration moyenne des plis vocaux dans les données concernées, cela représente une seule période ; cette valeur est comparable avec l'incertitude de la mesure elle-même (comme il a été expliqué au chapitre I, et ci-dessus au sujet des données anglaises) : la mesure de longueur se fait de la première fermeture glottique à la dernière fermeture glottique ; or la précision de la mesure elle-même est *à une période près*, non à quatre ou cinq millisecondes près (voir Lehiste 1970:10-17 ; Allen 1978). Dans les cinq cas où une tendance à l'allongement de la rime syllabique en condition de lecture *insistante* ressort de l'analyse statistique (aux trois tons chez M8, aux tons H et M chez M5), la différence moyenne (54 ms) est supérieure au seuil de perceptibilité (en anglais JND, Just-Noticeable-Difference) : « ... in the range of the durations of speech sounds—usually from 30 to about 300 msec—the just-noticeable-differences in duration are between 10 and 40 msec » (Lehiste 1970:13).

Les différences entre locuteurs et entre tons suggèrent qu'en naxi, l'énonciateur a une latitude entière dans l'emploi intonatif du paramètre de longueur.

locu-teur	ton	longueur moyenne (ms)		différence: cond. I moins cond. S		résultats du test-t sur la longueur (valeur de t, et p)	différence (dans les cas significatifs): condition I comparée à S
		cond. S	cond. I	en ms	en %		
F2	H	135	133	(=)	(=)	(0.22, 0.82)	
M5		269	330	+61	+23	s (-6.6, $6x10^{-9}$)	plus long
M7		256	250	(=)	(=)	(0.67, 0.50)	
M8		405	462	+57	+14	s (-4.9, $4x10^{-6}$)	plus long
M9		192	191	(=)	(=)	(0.13, 0.9)	
F2	M	147	146	(=)	(=)	(0.21, 0.83)	
M5		297	340	+43	+14.5	s (-4.7, 10^{-5})	plus long
M7		273	252	-21	-8.3	(1.8, 0.06)	
M8		423	480	+57	+13	s (-4.3, $4x10^{-5}$)	plus long
M9		219	209	-10	-4.8	(1.45, 0.15)	
F2	L	134	137	(=)	(=)	(-0.41, 0.68)	
M5		310	316	(=)	(=)	(-0.46, 0.64)	
M7		262	262	(=)	(=)	(0.04, 1)	
M8		423	475	+52	+12	s (-3.9, $2x10^{-4}$)	plus long
M9		226	213	-13	-6	(1.85, 0.07)	
moyenne		**265**	**280**	**+11**	**+5**		

Tableau 3.19. Longueur des syllabes naxi dans les deux conditions de lecture

2.5.5. L'intensité acoustique globale est plus élevée en lecture insistante, mais la différence est moindre que dans les données anglaises

Les différences constatées entre conditions sont moindres que celles observées dans les données anglaises (où elles étaient de l'ordre de 4 dB). Une intensité plus basse en condition de lecture *insistante* s'observe même chez le locuteur M8 au ton H, et chez le locuteur M5 au ton L.

Le nombre de types de voyelles dans les 120 syllabes du corpus est élevé : onze au plan phonémique, et douze au plan phonétique, du fait de la présence des deux allophones de /ɯ/ que sont [ɯ] et [z̩]. Les valeurs d'intensité ont été moyennées sur l'ensemble des voyelles, plutôt que de dresser une liste voyelle par voyelle.

locuteur	ton	intensité moyenne (dB), cond. S	intensité moyenne (dB), cond. I	différence: cond. I moins cond. S (dB)
F2	H	76.3	77.4	+1.1
M5		68.55	70.8	+2.25
M7		66.4	68.1	+1.7
M8		60.05	59.15	-0.9
M9		66.5	70.6	+4.1
F2	M	72.75	75.0	+2.25
M5		68.2	69.0	+0.8
M7		64.4	67.9	+3.5
M8		60.75	60.4	=
M9		64.3	70.0	+5.7
F2	L	69.5	70.7	+1.2
M5		65.6	63.0	-2.6
M7		61.1	64.1	+3
M8		56.1	56.15	=
M9		59.4	66.2	+6.8
MOYEN-NE		**Toutes conditions confondues : différence de 4.1 dB entre L et M, de 1.1 dB entre M et H**		**+1.9 dB**

Tableau 3.20. Intensité des syllabes naxi

2.5.6. Bilan des différences entre conditions de lecture dans les données naxi

En naxi, les stratégies d'insistance varient d'un locuteur à l'autre, et ne sont pas nécessairement les mêmes pour tous les tons chez un même locuteur. Elles reposent fréquemment sur un décalage de F_0 vers le haut, mais dans une plage limitée (ce qui, pour anticiper sur la section 4 ci-dessous, distingue nettement les données naxi des données anglaises). Chez les locuteurs M5, M7, F2, le rehaussement de la fréquence fondamentale est plus limité au ton L qu'aux tons M et H (et absent chez le locuteur M5) ; à l'opposé, M8 présente un léger rehaussement de fréquence fondamentale aux tons L et M, non au ton H. Des stratégies opposées s'observent d'un locuteur à l'autre dans l'emploi de la qualité de voix. Enfin, les deux premiers formants ont une légère tendance à être plus élevés en condition de lecture *insistante*, mais la variation est faible.

locuteur	ton	F_0 moyen	amplitude de la décroissance de F_0	longueur syllabique	quotient ouvert: partie de la syllabe sur laquelle la différence est significative		
					premier tiers	deuxième tiers	troisième tiers
F2	H	s	s		s		
M5		s		s	s	s	
M7		s			s	s	
M8			s	s			
M9		s	s				s
F2	M	s	s		s		
M5			s	s	s	s	
M7		s			s		
M8		s		s			
M9		s	s		s		
F2	L						
M5			s		s	s	
M7		s	s		s	s	s
M8		s		s		s	
M9		s	s		s	s	s
total des cas significatifs		**11**	**8**	**5**	10	7	3
					nb de cas où la diff. est significative sur plus des 2/3 de la syllabe : 6		

Tableau 3.21. Récapitulatif des différences entre conditions, données naxi. La différence de F_0 moyen est indiquée « s » si elle est significative sur plus de 75% de la syllabe. Ce seuil relativement élevé se justifie par le fait que cette mesure vise à refléter la plage de F_0, non les mouvements mélodiques.

Le tableau 3.21 peut se résumer en toutes lettres de la façon suivante : **le locuteur M5** produit les trois tons avec un quotient ouvert décroissant à mesure de la voyelle en lecture insistante, signe d'une qualité de voix qui devient plus *pressée* à mesure de la voyelle. Les autres paramètres varient selon le ton considéré : ses tons H et M sont allongés ; son ton L a une cible finale plus basse, son ton H une plage de F_0 légèrement plus élevée.

Le locuteur M7 recourt à une montée de plage de F_0 de l'ordre d'un ton de la condition *soignée* à la condition *insistante*. Dans les deux conditions de lecture, il réalise un contraste de hauteur très net entre les trois tons : il est frappant que le ton H, dans ses données, soit plus haut que chez la locutrice ; cela peut s'interpréter comme le signe d'une hyperarticulation.

Le locuteur M8 recourt à un allongement, et l'écart-type de F_0 devient plus bas en lecture insistante, ce qui peut être interprété comme le signe d'une précision accrue dans le contrôle

de la hauteur. En condition de lecture insistante, son ton M devient plus rectiligne ; la fin de son ton H présente un quotient ouvert significativement plus bas, indice d'une phonation *tenue :* l'évolution *non marquée* du quotient ouvert avant l'interruption de la phonation, bien attestée dans nos données, consiste en une montée finale, reflétant la décroissance progressive de la durée de la phase fermée, due à une adduction progressive des plis vocaux. Les courbes de F_0 du locuteur M9 sont globalement plus élevées, et plus planes (plus *tenues*), de sorte que les valeurs finales sont plus élevées en lecture insistante ; cette stratégie se rapproche de celle du locuteur M7 par la montée globale de plage de F_0 en lecture insistante, et de celle du locuteur M8 par le caractère plus rectiligne des courbes dans cette même condition.

La locutrice F2 présente un ton H nettement modifié (les courbes moyennées avec écarts-types ne se touchent pas, sauf en leur point terminal), un ton M moins fortement modifié, et un ton L faiblement modifié : la principale différence entre conditions pour le ton L est que quelques irrégularités finales de F_0 sont présentes en lecture soignée, dues au passage en mécanisme 0 (termes équivalents : laryngalisation, passage en voix craquée, en friture vocale ; en anglais : *laryngealisation, lapse into creaky voice, into vocal fry*) sur quelques items, phénomène qui accompagne fréquemment une phonation relâchée dans le bas de la tessiture du locuteur. Ces irrégularités disparaissent en lecture insistante, signe d'un contrôle plus précis de la phonation jusqu'à l'interruption du voisement.

3. Données vietnamiennes[1]

Rappel de l'enjeu scientifique de la prise en compte du vietnamien :

La prise en compte du vietnamien s'inscrit dans la perspective de l'**étude expérimentale des rôles prosodiques de la qualité de voix**, et tout particulièrement des relations entre fréquence fondamentale et qualité de voix. L'expérience menée sur le vietnamien, similaire à celles décrites ci-dessus pour l'anglais et le naxi, offre l'occasion d'une comparaison des résultats obtenus pour une langue à accent (l'anglais) et une langue à tons ponctuels (le naxi) avec des données équivalentes dans une langue dont les tons lexicaux combinent hauteur (spécifiée de façon **dynamique**, non en termes d'un simple niveau de hauteur) et qualité de voix.

Rappels concernant la langue vietnamienne : le vietnamien est une langue de la famille austro-asiatique, groupe môn-khmer, sous-groupe viet-muong. La variété étudiée est le parler de Hanoi, la capitale, lequel possède deux sous-systèmes tonals (l'un de six tons, en

[1] Les résultats présentés dans ce chapitre ont donné lieu à une communication de colloque (Michaud et Vu-Ngoc 2004) et un article de revue internationale (Michaud 2004a). Les dimensions du présent volume permettent d'exposer plus en détail les résultats.

syllabe finissant par une sonante ; l'autre de deux, en syllabe finissant par une occlusive). Une représentation schématique des courbes de fréquence fondamentale des tons, assortie d'indications sur la qualité de voix, est présentée sur la figure 4.1, au chapitre IV, section 3.1.5.3, où est également exposée l'origine diachronique des tons du vietnamien, qui est désormais bien connue dans ses grandes lignes (Maspero 1912 ; Haudricourt 1954, 1961 ; Ferlus 1998b, 2001).

3.1. Matériau linguistique

Les 126 mots vietnamiens choisis pour l'expérience illustrent trois tons (notés B2, D1 et D2, suivant en cela la notation étymologique de William Gedney [Gedney 1986, 1989], expliquée en détail au chapitre IV, section 3.1.5.3) qui contrastent par la hauteur et la qualité de voix, et paraissent suffire à fournir un aperçu des modifications intonatives des tons : parmi ces trois tons, **l'un comporte une constriction glottale finale (ton B2) ; le second est proche du premier par sa courbe de F_0, mais sans constriction glottale (ton D2) ; enfin, le troisième est un ton haut-montant (ton D1).**

Du fait de la complexité du système vocalique vietnamien, qui comprend des voyelles simples et des diphtongues ainsi que des oppositions de longueur, il a paru utile d'inclure dans le corpus l'ensemble des 14 noyaux vocaliques, afin de prendre en compte la part de variation *allotonique* qui est due aux voyelles[1]. Les syllabes retenues illustrent donc

- les 42 combinaisons de 14 noyaux vocaliques et des trois consonnes occlusives finales (/p/, /t/ et /k/), aux deux tons admis par ces syllabes : les tons D1 et D2 ; soit un total de 84 syllabes
- les 42 combinaisons de ces mêmes noyaux vocaliques suivis des consonnes nasales finales de même point d'articulation /m/, /n/ et /ŋ/, au ton B2.

Le tableau 3.22 résume ces combinaisons.

	ton D1	ton D2	ton B2
consonnes finales /p/, /t/, /k/	14 voyelles x 3 consonnes	14 voyelles x 3 consonnes	*combinaison non attestée*
consonnes finales /m/, /n/, /ŋ/	*combinaison non attestée*	*combinaison non attestée*	14 voyelles x 3 consonnes

Tableau 3.22. Tons et consonnes finales des mots du corpus vietnamien.

[1] Par inattention, ont été omis de la liste les noyaux vocaliques /wa/ et /wă/, dont la combinaison avec les deux séries de trois consonnes finales et avec les tons B2, D1 et D2 donnent les items suivants (en orthographe vietnamienne) : oạm, oạn, oạnh, oạp, oạt, oạc ; oặm, oặn, oặng, oặp, oặt, oặc ; oáp, oát, oác, oắp, oắt, oắc. Ces items figurent dans la liste complète employée lors de l'expérience de mesure de débit d'air (rapportée dans l'Annexe 2, section 2).

Trois des 126 syllabes ne sont pas des mots lexicaux : la voyelle /ɤ/ n'apparaît pas en combinaison avec les vélaires /k/ et /ŋ/ (en orthographe vietnamienne, ton inclus : *ớc, *ợc, *ợng). Ces syllabes ont néanmoins été incluses dans le corpus par souci d'exhaustivité. La lecture de *nonsense words* ne pose pas de difficulté particulière aux locuteurs rompus à la pratique de l'orthographe, comme l'ont observé Earle 1975:59 et Gsell 1980 pour le vietnamien, Thein Tun 1982:79 pour le birman ; au sujet de la rime *ơng du vietnamien, voir Haudricourt 1952:90, note 2, qui remarque : « … il suffit qu'elle soit virtuellement possible dans le système vietnamien ». La présence dans les données vietnamiennes de trois syllabes sans sens lexical ne nous paraît pas instaurer de différence importante avec les données naxi et anglaises.

Le tableau 3.23 présente les 126 syllabes enregistrées par les locuteurs M1, M2, M3 et F1 ; le numéro correspond à la position dans la liste lue.

27. áp 34. ạp 43. ạm	54. ắp 113. ặp 62. ặm	112. ấp 49. ập 61. ậm	68. ép 102. ẹp 77. ẹm	59. ếp 72. ệp 51. ệm	53. íp 48. ịp 74. ịm
13. át 107. ạt 4. ạn	29. ắt 40. ặt 90. ặn	92. ất 23. ật 41. ận	20. ét 87. ẹt 19. ẹn	24. ết 42. ệt 31. ện	26. ít 94. ịt 33. ịn
46. ác 37. ạc 98. ạng	2. ắc 3. ặc 103. ặng	104. ấc 1. ậc 79. ậng	45. ách 36. ạch 58. ạnh	6. ếch 12. ệch 108. ệnh	106. ích 18. ịch 8. ịnh
67. iếp 101. iệp 76. iệm	81. óp 14. ọp 80. ọm	100. ốp 66. ộp 75. ộm	56. ớp 15. ợp 5. ợm	52. úp 73. ụp 47. ụm	38. ứp 99. ựp 60. ựm
28. iết 21. iệt 89. iện	110. ót 84. ọt 85. ọn	22. ốt 30. ột 91. ộn	111. ớt 9. ợt 86. ợn	119. út 125. ụt 121. ụn	25. ứt 93. ựt 32. ựn
69. iếc 95. iệc 35. iệng	50. óc 97. ọc 71. ọng	78. ốc 11. ộc 105. ộng	115. *ớc 116. *ợc 117. *ợng	83. úc 82. ục 16. ụng	109. ức 7. ực 17. ựng
55. ướp 114. ượp 63. ượm	124. uốp 122. uộp 120. uộm				
10. ướt 88. ượt 39. ượn	126. uốt 123. uột 118. uộn				
44. ước 64. ược 57. ượng	96. uốc 65. uộc 70. uộng				

Tableau 3.23. Liste des items enregistrés par les locuteurs M1, M2, M3 et F1, en orthographe vietnamienne, arrangés par noyau vocalique. Le numéro renvoie à la position dans la liste en ordre aléatoire que lisent les locuteurs.

3.2. Réflexion sur les phrases-cadres

Les deux conditions de lecture sont les suivantes :

1) Lecture soignée : le consultant est invité à imaginer une situation où il enseigne la prononciation d'un mot à un enfant ou un étranger ne sachant pas lire le vietnamien. Il dit alors : « C'est le mot __ . » En vietnamien :

Đây	là	chữ	___.
DEICT.	COPULE	caractère	*mot-cible*

Chữ signifie littéralement *caractère*, le vietnamien adoptant en cela l'usage chinois. En chinois, chaque caractère correspond à la fois à un mot et à une syllabe[1] ; le vietnamien suit cet usage, et il est commun aujourd'hui encore de désigner comme « un caractère » chacune des syllabes, malgré la généralisation de l'écriture romanisée. (Au sujet de cette écriture, voir Haudricourt 1949, « L'origine des particularités de l'alphabet vietnamien », et Thompson 1965.)

[1] Voir ci-dessus note 2 en page 111.

2) Lecture insistante : le contexte est similaire au précédent, à ceci près que l'interlocuteur a déjà demandé à plusieurs reprises sans prendre soin de bien retenir la réponse qui lui était donnée. Cette condition de lecture demande à ce qu'une particule finale de discours soit ajoutée :

Đây	là	chữ	___	cơ mà !
DEICT.	COPULE	caractère	*mot-cible*	PARTICULE FINALE

« Mais enfin, c'est le mot __ ! »

L'élément final de phrase *cơ mà* exprime *l'insistance, l'affirmation forte*, avec une nuance de *contradiction* (Nguyên Nhu Y 1999). **En vietnamien, comme en chinois, en naxi et dans de très nombreuses autres langues d'Asie, il existe un large paradigme de particules finales de ce type, qui servent notamment à exprimer l'attitude de l'énonciateur**[1]. Dô Thê Dung, Trân Thien Huong et Georges Boulakia, étudiant le vietnamien, se rendent compte du fait en cours d'expérience : ils demandaient aux informateurs de lire une paire d'énoncés en passant de l'affirmation à l'interrogation par la seule intonation. Ceux parmi les sujets vietnamiens qui vivent en France depuis plusieurs mois s'exécutent, et appliquent le schéma d'interrogation du français, tandis que certains des consultants nouvellement arrivés de Hanoi refusent, disant que ces phrases-là n'existent pas (Dô Thê Dung, Trân Thien Huong *et al.* 1998:401).

Certes, la prise en compte des particules complique la tâche de l'enquêteur, puisqu'elle nuit à la symétrie des énoncés qu'il s'agit de comparer. Dans le cas présent, l'un des énoncés (celui qui correspond à la condition de lecture *insistante*) a deux syllabes de plus que l'autre, et place la syllabe-cible en position non finale, tandis qu'elle est finale en condition de lecture *soignée*. Une condition de contrôle a donc été ajoutée.

3) Condition de contrôle : ajout d'une troisième condition de lecture

La condition de contrôle est une variante de la condition soignée dans laquelle une particule finale exprimant le *respect dû à l'interlocuteur* a été ajoutée (cette particule s'écrit *ạ* en orthographe vietnamienne) :

Đây là chữ _____ ạ.

La consigne était : Vous vous adressez à une personne plus âgée que vous, qui a du mal à lire une syllabe. Cette condition est désignée ci-après comme condition de lecture *respectueuse*. Cette variante place la syllabe-cible en position non finale. Elle présente elle

[1] Au sujet des particules du vietnamien, voir l'ouvrage de référence de Thompson 1965. Parmi les descriptions des particules de langues apparentées au naxi, voir en particulier la description des particules du lahu par Matisoff 1973a, du lalo par Björverud 1998, du birman par Bernot 1980 et tout récemment par Vittrant 2004 ; concernant le chinois, langue dont la grammaire exerce une influence (directe ou indirecte) sur toute l'Asie du Sud-Est depuis les dynasties Qin et Han (vers les débuts de l'ère chrétienne), il existe de très nombreuses références, un grand classique étant l'ouvrage de Chao Yuen-ren 1968 ; voir également Marie-Claude Paris 1981 et références citées.

aussi un inconvénient : la syllabe *ạ* (/a/, au ton B2), n'ayant pas de consonne initiale, possède une attaque glottique (« attaque dure ») : le voisement commence après une occlusion laryngale (Thompson 1965:4), ce qui influence l'articulation laryngale du mot-cible, et par là la réalisation du ton dont ce mot est porteur. Cette dernière phrase-cadre n'a donc pas été utilisée en remplacement de la première, mais seulement comme une condition-contrôle, servant à vérifier qu'aucune différence flagrante n'apparaissait entre le cadre avec et sans particule finale *ạ*. Pour des raisons pratiques de dimensions du corpus, la condition de lecture *respectueuse* n'a été enregistrée que par le locuteur M4. Une partie des résultats correspondants sont présentés sur les figures 3.72b, 3.73b-c, 3.74b.

Les syllabes-cible ont été présentées une seule fois chacune, en ordre aléatoire, dans les deux conditions de lecture (*soignée* et *insistante*). Les instructions complètes en vietnamien sont les suivantes :

1) (anh dạy cho trẻ em hay người nước ngoài đang học tiếng Việt ; anh đọc một cách rất rõ)

Đây là chữ _____ .

2) (anh đã dạy cho con nhiều lần rồi, con vẫn chưa nhớ ; anh hơi bực tức, đọc một cách rất mạnh)

Đây là chữ _____ cơ mà !

3) (người cao tuổi không thấy rõ ; anh đọc cho người ấy một cách vừa rõ, vừa lịch sự.)

Đây là chữ _____ ạ.

Les consultants ont également lu un corpus plus étendu, auquel il est fait référence au cours de la discussion : un corpus de syllabes à sonante finale arrangées par séries tonales (tons A1 à C2) et par voyelles, hors phrase-cadre ; et un article de journal, qui permet de compléter les observations par la prise en compte de phénomènes de parole continue.

3.3. Bilan des enregistrements et résultats

Pour des raisons de commodité, les enregistrements n'ont pas été réalisés au Vietnam mais en France, au studio d'enregistrement du LIMSI-CNRS (Orsay). Les consultants sont trois locuteurs et une locutrice âgés de 19 à 24 ans, nés à Hanoi de parents hanoïens, et nouvellement arrivés de Hanoi (moins d'un an avant l'enregistrement). Ils ont été rétribués pour leur participation.

La locutrice F1, qui a participé la première aux enregistrements, a lu chacun des mots sous les deux conditions l'une à la suite de l'autre (premier mot en lecture soignée puis insistante, deuxième mot en lecture soignée puis insistante, etc.). Cette façon de lire est apparue assez difficile, du fait que tous les énoncés qui se succèdent doivent contraster de façon maximale par leur réalisation intonative. Par la suite, les locuteurs (M1 à M3) ont lu tous les mots en condition *soignée*, puis en condition *insistante*. (Les signaux se trouvent dans le dossier « sons vietnamiens ».)

La variation des formants n'a pas été étudiée sur le corpus considéré, pour une raison technique : la nasalisation d'une grande partie des voyelles (au sujet de l'effet acoustique du

couplage de la cavité nasale à la cavité orale, voir les rappels théoriques au chapitre I, section 2.1.1).

3.3.1. Analyse qualitative

L'analyse qualitative des données vise à caractériser la qualité de voix qui en vietnamien est associée à chacun des tons considérés, ce qui représente un préalable à l'analyse quantitative.

La forme du signal électroglottographique et de sa dérivée montre une nette différence entre le ton B2 d'une part, les tons D1 et D2 d'autre part. L'étude confirme ce que laissait entendre la description non instrumentale de Thompson 1965 : **les tons D1 et D2 (qui sont les deux tons pouvant apparaître sur les syllabes à occlusives finales) ne sont pas glottalisés.** Les figures 3.57 et 3.59 présentent des signaux typiques illustrant la constriction glottale finale du ton B2 et la fin de voisement non glottalisée du ton D2.

3.3.1.1. Ton B2

A la fin du ton B2 a lieu un allongement considérable de l'intervalle entre fermetures glottiques successives, cela pendant les trois à quatre derniers cycles. **L'inspection de signaux électroglottographiques et de leur dérivée indique une forte adduction des plis vocaux** pour le ton B2 : la phase fermée (pendant laquelle la glotte est fermée) est très longue. L'instant d'ouverture glottique apparaît sur la dérivée (sous la forme d'un pic négatif) peu avant le pic positif suivant (qui correspond à l'instant de fermeture suivant), comme l'illustre la figure 3.57. Cette observation rappelle quelque peu la voix craquée, décrite par Childers et Lee 1991:2394 comme une succession de « sharp, short pulses followed by a long closed glottal interval » (à ce sujet voir également Hollien 1974 ; Blomgren, Chen Yang *et al.* 1998), mais en voix craquée, la vibration des plis vocaux est irrégulière, tandis qu'à la fin du ton B2 l'allongement de la durée du cycle glottique a généralement lieu de façon monotone jusqu'à l'interruption du voisement (voir, à titre d'exemple, la figure 3.58, qui représente la laryngalisation finale au ton vietnamien C1, qui distingue ce ton de B2). **Cette caractéristique paraît motiver une distinction entre le phénomène de glottalisation observé au ton B2, qu'il paraît adéquat de décrire comme une *constriction glottale*, d'une part, et d'autre part la *laryngalisation*, qui est un passage en voix craquée.**

Certains cas intermédiaires ont néanmoins été constatés, qui ne se conforment pas dans le détail au modèle illustré par la figure 3.57. Dans certains cas, l'ouverture n'est pas très nette, et l'allure du signal électroglottographique et de son signal dérivé fait hésiter entre l'hypothèse d'une ouverture glottique juste après la fermeture, et l'hypothèse (qui s'impose pour la figure 3.57) d'une ouverture peu avant la fermeture suivante. Sur la figure 3.63, le signal électroglottographique peu après la fermeture est légèrement concave ; l'accélération

de la décroissance aux instants d'ouverture marqués par des flèches horizontales n'est pas nettement plus saillante que l'accélération constatée plus tôt dans le cycle glottique. Dans les cas comme celui-ci, lorsque le programme de détection donnait pour résultat une ouverture peu après la fermeture, et donc un quotient ouvert supérieur à 50% (souvent de l'ordre de 70%), notre choix a consisté à écarter ces valeurs comme physiologiquement non plausibles étant donné les valeurs de quotient ouvert des périodes qui précèdent[1].

Sur la figure 3.63, qui, comme la figure 3.64, comporte une portion concave dans le signal électroglottographique peu après la fermeture glottique au cours des premières périodes représentées (jusqu'à 160 ms), apparaît une petite fermeture glottique vers 210 ms. Son amplitude est réduite, et son incidence sur le signal audio à peine discernable. Il ne paraît pas prudent de proposer une interprétation de la portion de signal entre 170 et 210 ms, en y distinguant une phase ouverte et une phase fermée : une légère dépression est décelable juste avant la petite fermeture qui a lieu vers 210 ms ; dans la logique exposée ci-dessus, supposant que la constriction glottale se poursuive dans la droite ligne de ce qui précède, la glotte serait fermée jusqu'en ce point (à 200 ms) ; notre spéculation sur le mécanisme anatomique concerné serait la suivante : la glotte est fermée ; les plis vocaux ont commencé à se séparer dans leur partie inférieure aussitôt après la fermeture, mais demeurent accolés sur leur partie supérieure ; une bulle d'air s'échappe à 200 ms, occasionnant la légère dépression observée sur le signal électroglottographique, et entraînant aussitôt la petite fermeture observée. Cette conjecture amène à ne pas considérer le minimum local de la dérivée comme indice de l'instant d'ouverture dans un cas comme celui-ci, et à ne pas calculer de quotient ouvert pour ce cycle. Certains items au ton B2 présentent également des phénomènes irréguliers qui rappellent les observations que l'on peut faire sur les passages laryngalisés : des formes irrégulières de vibration, comme sur les figures 3.64 et 3.65 (à comparer à l'exemple typique de laryngalisation présenté en figure 3.58). Une compréhension affinée de ces phénomènes nécessitera d'autres techniques exploratoires : une analyse électromyographique offrirait un éclairage sur l'activité musculaire associée à la constriction laryngale, et sur la position temporelle de son relâchement ; des images des plis vocaux permettraient de connaître avec plus de précision la configuration glottique pendant ce passage glottalisé. La figure 3.66 illustre l'impossibilité de déterminer un instant

[1] Un argument *a posteriori* en faveur de cette procédure a été apporté en 2005 par Vu-Ngoc Tuân : confronté à la même observation, il observe que l'allure du signal audio plaiderait plutôt en faveur du second candidat au statut de pic d'ouverture, car elle tendrait à indiquer que le conduit vocal demeure fermé jusqu'au deuxième minimum local du signal dérivé : l'atténuation du signal (attribuable à l'ouverture de la glotte) intervient peu avant l'impulsion glottique suivante, ce qui correspond mieux au deuxième minimum détecté. Sur le spectrogramme (bas de la figure 3.65), il apparaît que les fréquences élevées demeurent nettes jusqu'à la dernière période, autre élément qui plaide contre l'idée d'un quotient ouvert élevé, celui-ci étant typiquement associé à une pente spectrale forte et donc à une atténuation de l'amplitude des harmoniques élevés. Il paraissait intéressant de rapporter cet élément de réflexion, bien qu'il ne soit pas déterminant.

d'ouverture après la dernière fermeture glottique au ton B2 ; en l'absence d'images de la glotte, il ne paraît pas possible de se prononcer sur la durée de la phase fermée. A la différence du ton D2 (figure 3.59), aucune oscillation périodique n'est observée après la dernière fermeture glottique.

Reprise du voisement après la constriction glottale dans environ 2 cas sur 5 en condition de lecture insistante

Lorsque la syllabe au ton B2 se trouve en position non finale, comme c'est le cas en condition *insistante*, une vibration peu régulière des plis vocaux reprend après la constriction glottale dans 37% des cas. La fréquence d'occurrence de ce phénomène, locuteur par locuteur, est : 66% pour F1, 2% pour M1, 76% pour M2, 41% pour M3. Il est possible que la différence entre locuteurs tienne au débit adopté par chacun d'eux, et au degré de netteté des frontières intonatives qu'ils placent après la syllabe-cible, paramètres qui n'étaient pas directement contrôlés lors de l'enregistrement. Des observations non quantifiées sur l'article de journal lu par les mêmes locuteurs (composante du corpus mentionné ci-dessus, section 3.2) suggèrent que cette reprise de voisement est plus fréquente en débit normal ou rapide.

En position finale (qui pour les données vietnamiennes correspond à la condition de lecture soignée ; voir la présentation des phrases-cadres en section 3.2), le voisement s'interrompt brutalement, le dernier cycle étant typiquement celui qui présente la durée la plus élevée et le quotient ouvert le plus bas. Le voisement ne reprend que dans un petit nombre de cas : moins de 5% pour M1, M2 et M3, 18% pour F1. Les figures 3.70b et 3.70c (locuteur M2) illustrent les caractéristiques des passages voisés qui succèdent à la constriction : leur quotient ouvert généralement élevé, et la variation forte d'un item à l'autre. Ce voisement succédant à la constriction glottale n'a pas été observé sur les syllabes ouvertes au ton B2 du corpus de syllabes enregistrées par séries par les mêmes locuteurs, seulement sur les syllabes à finales nasales. Dans les statistiques présentées ici, il a été choisi d'exclure la partie de « revoisement » : le moyennage des courbes sur l'ensemble de leur longueur n'aurait pas de sens, du fait du caractère non homogène des données moyennées. Une moyenne a en revanche été établie en ne prenant en compte que les périodes de voisement post-constriction ; elle donne une valeur de quotient ouvert de 56% (locuteurs confondus), ce qui confirme que la constriction glottale s'est alors relâchée.

Au plan perceptif, la rupture que constitue la constriction glottique est nette : pour une oreille non formée aux tons vietnamiens, les quelques cycles de voisement (cinq en moyenne) forment comme une prénasalisation ajoutée à la syllabe qui suit : un segment voisé nasal court mais audible est réalisé après la constriction glottique.

3.3.1.2. Ton D2

La qualité de voix au ton D2 est l'opposé de celle du ton B2, comme l'illustre la figure 3.59.
L'amplitude du signal électroglottographique augmente au cours de la deuxième moitié de la

rime, phénomène régulièrement observé lorsque la qualité de voix se rapproche de la voix soufflée *(breathy voice*, que Catford nomme *whispery voice* [Catford 1977 ; voir ci-dessus chapitre I, section 1.1.2]). (Cette figure illustre également le fait que l'amplitude du signal électroglottographique, qui reflète l'amplitude de la variation de la surface d'accolement des plis vocaux, ne correspond pas directement à l'amplitude du signal audio simultanément recueilli par le micro.) La vibration des plis vocaux disparaît progressivement : lorsque la dérivée devient presque plate, et qu'aucun pic de fermeture ni d'ouverture n'est plus visible, le signal électroglottographique montre encore quelques oscillations de très faible amplitude. Ces dernières oscillations semblent indiquer qu'**à l'interruption du voisement, l'adduction des plis vocaux cesse, la glotte est ouverte**, et les variations de surface d'accolement des plis vocaux deviennent très faibles (même si leur caractère périodique demeure décelable). Ce type d'interruption de voisement est comparable à celui qui est observé sur les rimes à sonante finale portant des tons non glottalisés (par exemple A1, A2 ou B1), catégories de syllabes qui figurent en grand nombre dans le corpus de séries de syllabes enregistré par les mêmes consultants (mentionné en section 3.2) : voir la figure 3.61, sur laquelle la périodicité visible sur le signal audio confirme que le changement d'allure du signal électroglottographique est dû à la diminution de la surface d'accolement des plis vocaux, non à la disparition de la périodicité du mouvement d'oscillation des plis vocaux. Néanmoins, les légères oscillations finales sont moins nombreuses sur les syllabes à occlusive finale : seules deux ou trois apparaissent, contre à peu près le double pour les rimes à finale sonante portant les tons A1, A2 ou B1. Cela peut s'interpréter comme une conséquence de l'augmentation de pression intra-orale qui suit l'occlusion (labiale, coronale ou palatale/vélaire selon les cas) : la pression transglottique diminue brutalement, ce qui interrompt plus rapidement la vibration des plis vocaux. Afin de préciser cette hypothèse, le haut de la figure 3.62 montre le spectrogramme correspondant à l'item représenté en figure 3.59, et repris en figure 3.60 avec indication de la position temporelle approximative de l'occlusion labiale (notée OL), déterminée d'après le spectrogramme. La figure 3.62 offre une comparaison avec le spectrogramme d'une syllabe à consonne finale de même point d'articulation, au ton B2 : l'espacement croissant des impulsions glottiques successives est nettement visible. Cela indique que les occlusives finales du vietnamien, notée phonologiquement comme non voisées (/p/, /t/, /k/) en l'absence d'opposition avec une autre série, sont phonétiquement voisées, au sens où le voisement se prolonge au-delà de l'occlusion orale. Ces consonnes pourraient donc aussi bien se noter comme des consonnes voisées, ce qui est la solution adoptée dans l'orthographe tibétaine : au plan phonétique, il n'y a pas de raison de postuler de différences entre les occlusives finales du vietnamien dans son état actuel et les occlusives finales du tibétain du VII[e] siècle, que Benedict 1972:13 postule être une série de « weakly articulated, imploded lenis surds »[1].

[1] Le terme « implosé » signifie ici « non relâché », et n'a pas le même sens que le terme « implosif » dans

Le signal électroglottographique montre l'absence de fermeture glottique immédiatement après la fin du voisement. Dans certains cas (2 sur 84 pour la locutrice F1, mais presque systématiquement chez le locuteur M3), deux ou trois pics positifs apparaissent sur le signal dérivé 60 à 250 ms après la fin du voisement ; ils sont d'amplitude plus faible que les pics de fermeture observés à l'intérieur d'un passage voisé. De tels mouvements d'ouverture et de fermeture de la glotte apparaissent également après les rimes à sonante finale portant un ton non glottalisé, ce qui indique qu'il n'y a pas lieu de les considérer comme caractéristiques du ton D2.

3.3.1.3. Ton D1

Les courbes de quotient ouvert du ton D1 présentent une certaine similarité avec celles du ton D2 : elles se trouvent dans la plage supérieure du locuteur, et s'élèvent au cours de la syllabe[1].

Le ton D1 est représenté sur les figures 3.69 (locuteur M1), 3.71d (locuteur M2), 3.74a-b (locuteur M3) et 3.77 (locutrice F1). **Rappelons qu'il s'agit d'un ton haut-montant, dont l'étude expérimentale vise à évaluer les modifications intonatives que connaît un ton dynamique.** L'expérience montre que le ton D1, de même que le ton D2, n'est pas glottalisé. Les enregistrements électroglottographiques de certaines des syllabes au ton D1 contiennent un artefact : pour 30 des 336 syllabes considérées (soit 9% des cas), la comparaison avec le signal audio montre que la vibration des plis vocaux cesse d'être enregistrée avant la fin de la syllabe. Cet artefact est peut-être dû à la rapide montée de F_0, qui s'accompagne d'une diminution de la surface vibrante des plis vocaux, ce qui rend plus difficile la mesure des

---

l'alphabet phonétique international, où il désigne les sons non pulmoniques produits en créant une sous-pression dans la bouche (consonnes également appelées *injectives* ou *claquantes*, voir Haudricourt 1950:172-173). Ainsi, les sons /ɓ/ et /ɗ/ du vietnamien (les b et đ de l'orthographe vietnamienne romanisée) sont notés en API comme des consonnes *implosives*. Il peut pourtant être indiqué de les décrire comme *préglottalisées* : l'attaque glottique d'une préglottalisée diffère de celle d'une sonore de la même façon qu'une voyelle à attaque dure diffère d'une voyelle à attaque douce, et ne s'accompagne pas nécessairement du geste caractéristique des implosives : une descente du larynx produisant une dépression dans l'espace clos en bas par l'occlusion laryngale et en haut par une occlusion buccale (Haudricourt 1950, « Les consonnes préglottalisées en Indochine »).

[1] Une comparaison détaillée des tons D1 et D2 montre que les valeurs de quotient ouvert du ton D1 ont tendance à être plus basses pour les locuteurs masculins (dans les deux conditions de lecture) : les moyennes sont les suivantes : M1: - 12.5%; M2: - 4.80%; M3: - 13.8%. Ce phénomène paraît explicable en termes de polarisation de qualité de voix : les tons B2 et D2 sur les rimes ayant une consonne finale de même lieu d'articulation présentent une forte similarité phonétique, du fait que les nasales finales des rimes au ton B2 sont peu audibles ; ces deux types de rimes se différencient principalement par la qualité de voix ; ils se trouvent donc dans une relation d'opposition, d'où la qualité de voix relativement *détendue* de D2 (à l'opposé de la glottalisation de B2), tandis que D1, qui n'entre pas dans ce type d'opposition du fait que les contrastes sont suffisamment assurés présente des valeurs de quotient ouvert moins extrêmes. Une discussion détaillée à ce sujet est proposée dans Michaud 2004a.

fluctuations de surface de contact des plis vocaux par l'électroglottographe. Dans ces cas-là, la syllabe entière a été exclue des calculs ; mais la présence de cet artefact dans une partie des données jette un doute sur la précision de la mesure en fin de syllabe dans les autres cas. Une investigation fine des changements de configuration laryngale qui ont lieu à la fin du ton D1 devra recourir à d'autres techniques exploratoires.

Chez la locutrice F1, les valeurs de quotient ouvert des tons D1 et D2 sont proches, mais ne sont pas directement comparables : en effet, F1 produit les fréquences fondamentales basses telles que celle du ton D2 en « voix de poitrine » (plus précisément appelé *mécanisme 1* ; au sujet des mécanismes phonatoires, voir l'entrée correspondante du Glossaire), et les fréquences fondamentales plus élevées, par exemple en fin de ton D1, en « voix de tête » (plus précisément appelé *mécanisme 2*, mécanisme dans lequel le quotient ouvert est systématiquement élevé et connaît moins de variation qu'en mécanisme laryngien 1).

La comparaison des réalisations du ton D1 en fonction des conditions de lecture n'a pas été poussée aussi loin que pour les tons B2 et D2, du fait des incertitudes concernant les mesures électroglottographiques sur ce ton et du fait que les valeurs de quotient ouvert ne sont pas directement comparables d'un mécanisme laryngien à l'autre. **Il ressort néanmoins clairement que les locuteurs M1 et M2 recourent à une élévation globale de la F_0, et le locuteur M3 à un abrègement de la durée syllabique.**

3.3.2. Analyse quantitative

L'analyse quantitative qui suit se concentre essentiellement sur les tons B2 et D2, à la fois pour la raison technique (exposée ci-dessus) d'une incertitude concernant la mesure des derniers cycles glottiques au ton D1, et du fait de l'intérêt particulier que présente l'opposition de qualité de voix entre les tons B2 et D2 dans la perspective d'une étude des relations entre emploi lexical et emploi intonatif de la qualité de voix.

3.3.2.1. Longueur des cycles glottiques (fréquence fondamentale) : description du contraste entre tons et entre conditions de lecture

Pour tous les locuteurs, vers la fin du ton B2, la longueur des cycles augmente nettement. Des courbes représentatives de F_0 (pour le locuteur M1) sont présentées sur les figures 3.67a-b et 3.68a-b. Les tableaux 3.24 et 3.25 présentent les valeurs moyennes sur la syllabe (voir également le graphique 3.7, page 492) et l'évolution de la fréquence fondamentale au cours de la syllabe, respectivement. La différence moyenne entre les deux derniers cycles est de 52% (locuteurs et conditions de lecture confondus), ce qui revient à dire que la longueur du cycle augmente de moitié. Il demeure néanmoins possible de calculer l'inverse de la durée qui sépare les fermetures glottiques successives, suivant en cela la recommandation d'Osamu Fujimura (*in* Thongkum 1988:333). A ce point de la courbe mélodique, il n'y a plus à proprement parler de période, puisque cette notion suppose une forte similarité entre

cycles successifs. Une méthode de détection de la fréquence fondamentale par corrélation (tel que l'algorithme DECOM défini par Nathalie Henrich pour le signal électroglottographique [Henrich, d'Alessandro *et al.* 2004]), ne peut donc s'appliquer à cette portion de signal. Lorsque le ton B2 vietnamien est réalisé de façon canonique (comme c'est le cas dans les données étudiées ici), chaque cycle glottique est généralement bien défini, à la différence de la voix laryngalisée, phonation « chaotique » selon l'expression de Philippe Martin (communication personnelle), où le passage d'une bulle d'air entre les plis vocaux accolés semble se faire d'une façon relativement aléatoire. Par commodité, nous retenons donc pour la description du ton B2 le terme de « fréquence fondamentale », F_0, pour l'inverse de la durée du cycle, quel que soit le degré de périodicité du signal.

Au cours du ton D2, la fréquence fondamentale descend graduellement, descente qui s'accentue légèrement en fin de rime. La différence moyenne entre le point le plus haut et le point le plus bas de la syllabe (locuteurs et conditions de lecture confondus) est de 26%, soit environ 2 tons musicaux (contre 64% pour le ton B2).

locu-teur	ton/ type de syllabe	F_0 moyen (Hz) et écart-type		différence : cond. I moins cond. S, en % (gauche) et en tons musicaux (droite)		partie de la syllabe (sur les 100 échantillons également espacés dans le temps) sur laquelle la différence est significative	différence d'écart-type de F_0 : cond. I comparée à S
		cond. S	cond. I				
F1	ton B2 (finale nasale)	228 (12.9)	240 (25.7)	+5.4	½	38e au 97e	+64%
M1		129 (10.0)	168 (12.2)	+30	2 ½	1er au 100e	+13%
M2		115 (9.4)	126 (14.1)	+9.8	¾	27e au 98e	+71%
M3		128 (9.2)	131 (15.7)	+2	< ¼	3e au 21e (plus bas en cond. I) et 52e au 100e (plus haut en cond. I)	+64%
F1	ton D2 (finale occlusive)	224 (8.5)	248 (12.7)	+11	proche de 1	1er au 96e	+37%
M1		126 (4.5)	144 (4.8)	+15	1 ¼	1er au 100e	+7%
M2		107 (6.0)	131 (9.7)	+22	1 ¾	1er au 99e	+56%
M3		121 (5.4)	124 (5.9)	+2	< ¼	72e au 100e	-16%

Tableau 3.24. Statistiques sur les valeurs moyennes de F_0 des données vietnamiennes. Chaque valeur représente la moyenne d'un ensemble de 42 syllabes de même catégorie tonale.

On relèvera ici

- l'écart-type globalement plus élevé en condition de lecture insistante, une plus forte variabilité figurant parmi les corrélats de l'énonciation *insistante* chez ces locuteurs
- l'élévation de plage de F_0 en lecture insistante chez tous les locuteurs, à des degrés variés.

Le tableau 3.25a présente l'évolution de F_0 au cours de la syllabe, selon le même principe que pour les données naxi et anglaises : après exclusion du premier cycle glottique et du dernier, pour chaque courbe.

locuteur	ton/ type de syllabe	décroissance de F_0 (en tons musicaux) du 2e à l'avant-dernier cycle glottique		résultat du test-t sur la différence d'évolution de F_0 entre cond. S et cond. I (entre parenthèses : valeur de t et valeur de p)	différence d'évolution (dans les cas significatifs): condition I comparée à S
		Lecture soignée	Lecture insistante		
F1	ton B2 (finale nasale)	-2.7	-4.0	s (t_{38} = 3.15, p = 0.002)	décroissance plus forte
M1		-2.7	-2.6	(t_{41} = 1.74, p = 0.08)	
M2		-1.3	-3.9	s (t_{41} = 8.54, p = 6 x 10^{-13})	décroissance plus forte
M3		-1.9	-2.8	s (t_{41} = 3.45, p = 0.0009)	décroissance plus forte
F1	ton D2 (finale occlusive)	-1.4	-0.8	s (t_{38} = -2.58, p = 0.01)	décroissance moins forte
M1		-1.1	-1.3	s (t_{41} = 3.1, p = 0.003)	décroissance plus forte
M2		-0.9	-0.9	(t_{41} = 0.82, p = 0.42)	
M3		-1.0	-1.4	(t_{41} = 2.5, p = 0.01)	décroissance plus forte

Tableau 3.25a. Statistiques sur la variation de F_0 du deuxième à l'avant-dernier cycle glottique. Chaque valeur représente la moyenne d'un ensemble de 42 syllabes de même catégorie tonale. La lettre s signifie « significatif ».

Une mesure complémentaire a également été tentée : entre le point le plus haut et le point le plus bas de chaque courbe ; cette mesure, réalisée uniquement pour les données vietnamiennes, visait à faire ressortir le fort degré d'allongement des cycles glottiques d'un cycle à l'autre en fin de ton B2, après un début de courbe parfois montant, comme il apparaît sur les figures 3.67, 3.70, 3.72 et 3.75.

locuteur	ton/ type de syllabe	décroissance de F_0 (en tons musicaux)		résultat du test-t sur la décroissance de F_0 (entre parenthèses : valeur de t_{41} et valeur de p)	Proportion de la durée totale de la syllabe sur laquelle a lieu la décroissance de F_0 (en %)		résultat du test-t sur la durée de la décroissance de F_0 (entre parenthèses : valeur de t_{41} et valeur de p)
		Lecture soignée	Lecture insistante		Lecture soignée	Lecture insistante	
F1	ton B2 (finale nasale)	2.30	1.95	s (-4.87; 2*10^{-5})	77	86	s (2.09; p > 0.03)
M1		2.26	2.30	(-1.50; 0.14)	67	63	(0.21; p > 0.5)
M2		1.93	2.42	s (-13.85; < 10^{-9})	69	62	s (3.82; p < 10^{-3})
M3		2.11	2.09	s (-3.50; 0.001)	62	75	(0.69; p > 0.4)
F1	ton D2 (finale occlusive)	5.18	6.20	(1.42; 0.16)	72	63	(0.16; p > 0.5)
M1		5.0	4.6	(-1.81; 0.07)	61	59	(-0.07; p > 0.5)
M2		4.0	7.1	s (-4.85; 2*10^{-5})	70	41	s (3.76; p < 10^{-3})
M3		5	5.8	(-0.25; 0.80)	53	48	(0.36; p > 0.5)

Tableau 3.25b. Statistiques sur la différence entre les valeurs les plus élevées et les plus basses de F_0 sur la syllabe. Chaque valeur représente la moyenne d'un ensemble de 42 syllabes de même catégorie tonale. La lettre s signifie « significatif ».

Ce n'est que chez le locuteur M2 qu'une décroissance de F_0 significativement plus forte a lieu en lecture insistante pour les deux tons ; décroissance qui est significativement plus rapide (s'étendant sur une proportion plus réduite de la rime).

3.3.2.2. Quotient ouvert : les valeurs des tons B2 et D2 contrastent de façon maximale ; elles varient peu en fonction de la condition de lecture

Les valeurs de quotient ouvert des tons B2 et D2 sont tracées au bas des figures 3.67 et 3.68 ; les valeurs chiffrées sont présentées dans les tableaux 3.26 et 3.27. **Elles contrastent entre elles de façon maximale.** Rappelons que les valeurs sont à rapporter à la plage de variation donnée par Titze 1988:236, qui s'étend de 30% (voix pressée, avec forte adduction des plis vocaux) à 70% (voix soufflée : *breathy voice*).

En ce qui concerne les rimes au ton B2, le quotient ouvert est dans la plage inférieure du locuteur sur toute la syllabe : la voix est *pressée*, et la compression s'accroît au cours de la seconde partie de la rime, terminant par des valeurs très basses. Ce ton présente fréquemment des valeurs finales de quotient ouvert inférieures à 30% pour le dernier cycle glottique : celles-ci sont caractéristiques d'une constriction forte, incompatible avec la poursuite du voisement ; des valeurs aussi basses ont de quoi susciter la surprise de chercheurs qui travaillent essentiellement sur la voix dite *modale*[1]. La figure 3.67a montre qu'il n'y a pas d'exception à cette observation chez le locuteur M1 (voir également les figures 3.70, 3.72, 3.75 ; la reprise du voisement après la constriction glottique qui apparaît dans la figure 3.70b-c est décrite plus bas).

En ce qui concerne les rimes au ton D2 (bas des figures 3.68, 3.71, 3.73, 3.76), les valeurs de quotient ouvert sont dans le milieu de la plage du locuteur. Elles augmentent au cours de la rime, ce qui indique une décroissance du degré d'adduction des plis vocaux.

Les plages de variation des tons B2 et D2 sont pratiquement sans contact l'une avec l'autre, mis à part les valeurs dispersées qui s'observent en début de syllabe du fait de l'attaque glottique de la rime (systématique en vietnamien en l'absence de consonne initiale). Au ton D2, la valeur moyenne sur la dernière période (locuteurs et conditions confondues) est légèrement supérieure à 60%, valeur qui suggère un débit d'air élevé (voir Rothenberg et Mahshie 1988 et la discussion au sujet de l'interprétation du quotient ouvert présentée au chapitre I, section 1.2.3.3). Ces valeurs nous paraissent refléter le fait que la qualité de voix est spécifiée lexicalement de façon précise pour ces tons vietnamiens.

Les valeurs observées témoignent d'une remarquable constance de la qualité de voix d'une condition de lecture à l'autre, et d'une non moins remarquable uniformité dans la direction que prend la légère variation observée : le quotient ouvert moyen est bas au ton B2, et plus bas en lecture insistante qu'en lecture soignée ; il est relativement élevé au ton D2, et plus élevé en lecture insistante qu'en lecture soignée (voir le tableau 3.26). La tendance à une plus forte variabilité en condition de lecture insistante qui s'observait pour la fréquence fondamentale n'apparaît pas pour le quotient ouvert.

---

[1] Ainsi, de telles valeurs n'ont jamais été rencontrées par Nathalie Henrich, dont la thèse (Henrich 2001) porte essentiellement sur la voix lyrique occidentale.

locuteur	ton/ type de syllabe	quotient ouvert moyen (%)		différence : cond. I moins cond. S, en %	partie de la syllabe (sur les 100 échantillons également espacés dans le temps) sur laquelle la différence est significative	écart-type de cond. I comparé à S
		cond. S	cond. I			
F1	ton B2 (finale nasale)	50	47	-4.6%	69 à 100	+8%
M1		32	32	-0.3%	38 à 48	-6%
M2		43	38	-11%	57 à 100	+13%
M3		38	37	-2.9%	77 à 100	+27%
F1	ton D2 (finale occlusive)	63	64	+1.8%	80 à 100	-30%
M1		51	53	+5.6%	7 à 96	+16%
M2		53	56	+5.6%	55 à 100	+20%
M3		57	60	+6.0%	8 à 100	-18%

Tableau 3.26. Statistiques sur les valeurs de quotient ouvert. Chaque valeur représente la moyenne d'un ensemble de 42 syllabes de même catégorie tonale.

3.3.2.3. Stratégies opposées dans l'emploi intonatif du paramètre de longueur

Les **stratégies opposées d'un locuteur à l'autre (et d'un ton à l'autre)** dans l'emploi intonatif du paramètre de longueur sont commentées dans le bilan des variations entre condition de lecture soignée et condition de lecture insistante en section 3.3.3.

locuteur	ton/ type de syllabe	longueur moyenne (ms)		différence: cond. I moins cond. S		résultat du test-t sur la longueur syllabique (entre parenthèses : valeur de t_{41} et valeur de p)
		cond. S	cond. I	en ms	en %	
F1	ton B2 (finale nasale)	126	143	+17	+13%	(-1.94 ; $p > 0.05$)
M1		145	149	(=)	(=)	(-1.26; $p > 0.2$)
M2		126	210	+84	+67%	s (13.0; $p < 10^{-12}$)
M3		166	183	+17	+10%	(-1.61; $p > 0.1$)
F1	ton D2 (finale occlusive)	121	107	-14	-12%	s (7.24; $p < 10^{-7}$)
M1		136	144	+8	+6%	s (-3.08; $p < 0.005$)
M2		131	139	+8	+6%	(-1.13; $p > 0.2$)
M3		145	119	-26	-18%	s (10.4; $p < 10^{-12}$)

Tableau 3.27. Statistiques sur la durée syllabique. Chaque valeur représente la moyenne d'un ensemble de 42 syllabes de même catégorie tonale. La lettre s signifie « significatif ».

Les mises en garde exprimées ci-dessus (section 1.7.4, dans le commentaire des données anglaises) au sujet des mesures de longueur s'appliquent également dans l'interprétation des valeurs du tableau 3.27 : la variation moyenne de longueur est faible (mis à part pour le ton B2 du locuteur M2), rapportée à la fréquence de vibration moyenne des plis vocaux dans les données concernées. Cette valeur est comparable avec l'incertitude de la mesure elle-même : la précision de la mesure de longueur est *à une période près* (voir Lehiste 1970:10-17 ; Allen 1978, « Vowel duration measurement: a reliability study »).

3.3.2.4. Intensité acoustique globale (mesure réalisée sur une partie du corpus seulement)

La différence d'intensité entre les deux conditions de lecture est de 5.8 dB en moyenne pour le locuteur M1 (voir le tableau 3.28). Pour des raisons techniques, les mesures d'intensité acoustique globale n'ont pas été réalisées systématiquement, seulement pour les données du locuteur M1 ; les observations concernant les trois autres locuteurs ont consisté en relevés manuels sur une dizaine de paires d'items pour chaque locuteur. De même que pour la fréquence fondamentale, le locuteur M3 paraît se distinguer des locuteurs F1, M1 et M2 : chez le premier, il ne paraît pas y avoir d'augmentation significative d'intensité (pas plus que de F_0) ; chez les seconds, une forte augmentation d'intensité s'observe en lecture insistante.

ton	intensité moyenne (valeurs négatives en décibels, par rapport à 0 dB)		différence (condition I - condition S) en décibels
	condition S	condition I	
B2	-28.44	-22.69	5.75
D1	-24.32	-19.17	5.15
D2	-28.69	-22.04	6.65
moyenne			5.85

Tableau 3.28. Intensité acoustique globale des syllabes lues par le locuteur vietnamien M1.

Il existe des travaux relativement abondants sur les seuils psycho-acoustiques et psycho-linguistiques de perception d'une différence d'intensité. Rappelons simplement que le seuil différentiel d'intensité pour la parole est d'environ 3 dB (Rossi 1999:206).

3.3.2.5. L'hypothèse d'une corrélation entre DECPA et débit d'air

DECPA (Derivative-Electroglottographic Closure Peak Amplitude ; voir l'Annexe 2, section 3) **fournit une indication supplémentaire qui paraît corrélée avec le débit d'air, mais ne nous paraît pas apporter d'élément décisif.** Les figures 3.78 et 3.79 indiquent les valeurs moyennes pour les tons D2 et B2, respectivement, selon le même principe de simulation syllabique utilisé pour F_0 et le quotient ouvert. Les valeurs absolues ne signifient rien en elles-mêmes, car elles dépendent du gain choisi à l'enregistrement ; en revanche, ces valeurs peuvent être comparées d'un ton à l'autre et d'une condition à l'autre. Les tracés montrent une différence importante (quoique très variable d'un item à l'autre) entre ces deux tons, et d'une condition de lecture à l'autre : la courbe du ton B2 forme un arc de cercle ; en condition de lecture *insistante*, les points de départ et d'arrivée sont inchangés, tandis que la courbe devient légèrement plus convexe ; la courbe du ton D2 est montante, montée qui s'accentue sur la deuxième moitié de la rime en lecture insistante. Une certaine similarité apparaît entre ces courbes de DECPA et les courbes de débit d'air du locuteur M4 (Annexe 2, section 2). Cela suggère une utilisation possible du paramètre DECPA : il permettrait de formuler des hypothèses concernant le débit d'air, DECPA étant corrélé au débit d'air. Le degré de corrélation entre ces deux paramètres n'a pas été étudié ici de façon statistique

faute d'une quantité suffisante de données de débit d'air et de la qualité non optimale de l'enregistrement électroglottographique réalisé simultanément avec ces mesures de débit d'air.

3.3.3. Bilan : les caractéristiques de qualité de voix ressortent plus nettement encore en lecture insistante (constriction glottale du ton B2, phonation modale des tons D1 et D2)

Rappelons que la question linguistique qui est en jeu est celle du degré de variation intonative des paramètres de hauteur, de qualité de voix et de longueur, sachant que tous trois ont un emploi lexical : hauteur et qualité de voix définissent les tons ; la longueur joue un rôle dans des oppositions vocaliques.

Le quotient ouvert varie peu d'une condition à l'autre ; en revanche, tous les locuteurs présentent des valeurs de quotient ouvert qui sont stables (pas de différence significative entre conditions) ou vont dans le sens d'une réalisation plus extrême de la qualité de voix lexicalement associée au ton : quotient ouvert plus bas pour le ton B2, signe d'une constriction glottale plus forte ; quotient ouvert plus haut pour le ton D2, signe d'une adduction moins forte des plis vocaux. Il paraît s'agir d'une forme d'**hyper-articulation de la qualité de voix lexicale**. (Les valeurs de quotient ouvert pour le ton D1 n'ont pas été soumises à traitement statistique, du fait des incertitudes techniques rapportées en section 3.3.1.3.) Les courbes correspondant aux deux conditions de lecture, en bas des figures 3.67c et 3.68c, sont très proches l'une de l'autre ; voir également les figures 3.70c, 3.71c, 3.72a, 3.75 et 3.76. Lorsqu'une différence significative est constatée, elle porte généralement sur la deuxième partie de la rime (voir détails dans le tableau 3.26). **Le degré de variation du quotient ouvert entre conditions est plus faible que le degré de variation de fréquence fondamentale et d'autres paramètres.**

En revanche, hauteur, longueur et intensité varient selon des stratégies qui diffèrent d'un locuteur à l'autre. Les paramètres qui varient en fonction de la condition de lecture diffèrent d'un locuteur à l'autre et, chez certains locuteurs, d'un ton à l'autre. La distribution des données autorisant le recours à des tests paramétriques, la comparaison statistique des deux conditions consiste en tests-t *(two-tail)* appliqués à chacun des points (équidistants dans le temps) des courbes rééchantillonnées. Comme précédemment, des tests non appariés sont employés dans les cas où l'absence de certaines valeurs de quotient ouvert a pour effet des groupes de taille différente. **La fréquence fondamentale est significativement différente sous insistance dans la majorité des cas ; la tendance générale est à une élévation de la plage de réalisation du ton.** Chez le locuteur M1, la plage de fréquence fondamentale change ; chez le locuteur M2, l'amplitude et la durée de la diminution de F_0 changent de façon conséquente ; chez les locuteurs F1 et M3, les paramètres affectés dépendent du ton. Les différences de longueur entre conditions de lecture varient en fonction du locuteur et du ton lexical concerné : significativement plus court sous insistance (ton D2, chez deux des locuteurs), ou

plus long (ton D2 chez un locuteur, ton B2 chez un autre), ou sans différence significative (dans les quatre autres cas). Un bilan est proposé dans le tableau 3.29.

locuteur	ton	F_0 moyen	évolution de F_0 du 2e à l'avant-dernier cycle	baisse de F_0 du point le plus élevé au plus bas	longueur syllabique	quotient ouvert : partie de la syllabe sur laquelle la différence est significative
F1	B2		s			moins d'1/3
	D2	s	s	hs	hs	moins d'1/3
M1	B2	hs		s		moins d'1/3
	D2	hs	s		s	environ 90%
M2	B2	s	hs	hs	hs	environ 50%
	D2	s		s		environ 50%
M3	B2		s			moins d'1/3
	D2		s	s	hs	environ 90%
nb de cas significatifs		5	6	5	4	2

Tableau 3.29. Bilan des différences entre condition soignée et condition insistante ; hs : « hautement significatif », $p < 10^{-4}$; s: « significatif », $p < 0.05$; blanc: non significatif, $p > 0.05$.

4. Synthèse des résultats : comparaison entre les trois langues

% de cas significatifs par langue	F_0 moyen	amplitude de la baisse de F_0	longueur syllabique	quotient ouvert	F_1	F_2	F_3	intensité acoustique globale
anglais	**80**	**30**	**60**	**10**	**13**	**23**	**26**	**70**
naxi	**73**	**53**	**33**	**40**				
vietnamien	**63**	**75**	**50**	**25**				

Tableau 3.30. Bilan simplifié des différences entre conditions de lecture dans les trois langues.

4.1. Importance des modifications locales

Le tableau 3.30 (dont une synthèse visuelle est proposée sur le graphique 3.10, en bas à gauche de la page 493) fait ressortir les faits suivants :

- **Dans les trois langues considérées, des modifications statistiquement significatives des mots-cible ont lieu d'une condition de lecture à l'autre.**
- **Les modifications subies par les courbes de F_0 sont globalement de moindre amplitude dans les langues à tons étudiées qu'en anglais.**

Chez le locuteur naxi M8, la hauteur n'est pas affectée de façon significative par le changement de condition de lecture. L'augmentation d'intensité est également un peu plus faible en naxi qu'en anglais (environ 2 dB, contre 4 chez les locuteurs anglais). Néanmoins, il apparaît que le lien entre F_0 et intensité n'est pas le même dans les trois langues. Le paragraphe suivant s'arrête sur la question de la relation de ces deux paramètres : à niveau

de mise en valeur comparable, l'intensité est comparativement plus modifiée en naxi qu'en anglais. **Les faits observés conduisent à une hypothèse sur le *contrôle du paramètre de F_0* qui serait exercé par les locuteurs de langues à tons.**

4.2. Relation entre fréquence fondamentale et intensité dans les trois langues étudiées

Au plan linguistique, il paraît intéressant de comparer la corrélation entre F_0 et intensité vocale dans les trois langues étudiées, pour voir si les langues présentent, de façon tendancielle, une relation différente entre F_0 et intensité selon qu'elles possèdent ou non des tons lexicaux. Il est connu que, toutes choses égales par ailleurs, fréquence fondamentale et intensité acoustique globale sont corrélées (voir par exemple Gramming, Sundberg *et al.* 1988, Titze et Sundberg 1992, Liénard et Di Benedetto 1999). L'expérience du chant montre néanmoins qu'il est possible de maintenir l'un des deux paramètres relativement constant tout en modifiant l'autre. Dans le chant lyrique occidental, le chanteur apprend à décorréler hauteur et intensité ; l'un des exercices consiste à modifier l'intensité en gardant la hauteur constante (voir par exemple les mesures rapportées par Henrich 2001:140).

4.2.1. En anglais : fortes modifications de F_0 et de l'intensité ; en naxi et vietnamien : modification relativement plus élevée de l'intensité

Sur les monosyllabes anglais, le changement d'intensité acoustique globale entre énonciation soignée et énonciation insistante est en moyenne de 2.70 dB (locuteurs confondus) ; la différence de F_0 moyen est de 1.8 tons musicaux. **De ces deux valeurs peut se déduire une approximation de la relation qui unit les deux paramètres : 1.5 dB pour 1 ton musical.** Dans cette approximation grossière, la relation entre les deux paramètres est considérée comme linéaire, ce qui est bien sûr une simplification[1], mais suffira ici pour le propos de la comparaison entre langues. Les syllabes non accentuées ne sont pas retenues car elles représentent un ensemble trop différent des monosyllabes du naxi et du vietnamien.

Le tableau 3.31 présente ces mêmes données pour les deux types de syllabes accentuées du corpus anglais (monosyllabes, et première syllabe de disyllabe), et leur moyenne, en vue de la comparaison avec les données naxi et vietnamiennes.

[1] En particulier, cette moyenne ignore le découpage de la tessiture de chaque locuteur en **mécanismes phonatoires**.

langue	type de syllabe	ΔF_0 : différence entre condition I et S, en tons musicaux	ΔI : différence entre condition I et condition S, en décibels	$\Delta I / \Delta F_0$: nombre de dB pour 1 ton musical
anglais	monosyllabe	1.78	2.7	1.5
	1e syll. de disyll.	1.98	4.2	2.1
	moyenne	1.88	3.45	1.8
naxi	tons H, M et L confondus	0.7	1.9	2.7
vietnamien	tons B2, D1 et D2 confondus	1.9	5.8	3

Tableau 3.31. Rapport entre variation de F_0 et variation d'intensité acoustique globale dans les données des trois langues étudiées. Les valeurs données pour le vietnamien représentent la moyenne de 252 syllabes (1 locuteur), contre 1200 syllabes pour le naxi (5 locuteurs) et 240 pour l'anglais (5 locuteurs).

En vietnamien, l'augmentation de F_0 est de 1.9 ton musical, soit une valeur comparable à la moyenne des locuteurs anglais. En revanche, l'augmentation d'intensité est nettement supérieure. En naxi, la variation est globalement plus faible : la variation d'intensité est presque deux fois plus faible qu'en anglais, la variation de hauteur presque trois fois plus faible. La relation entre variation d'intensité et de F_0 lors du passage de la condition soignée à la condition insistante est du même ordre de grandeur en naxi et en vietnamien (de l'ordre de 3 dB pour un ton musical) ; elle est supérieure à ce qui est observé pour l'anglais (relation de l'ordre de 2 dB pour un ton musical).

4.2.2. Le rapport $\Delta I / \Delta F_0$ est plus bas lors d'un changement de ton lexical que lors d'un changement de lecture soignée à lecture insistante

Les valeurs calculées ci-dessus sont d'autant plus parlantes lorsqu'elles sont rapportées aux différences entre les trois tons lexicaux du naxi en termes de F_0 et d'intensité, présentées dans le tableau 3.32. Le graphique 3.11 (p. 493) offre une synthèse des tableaux 3.31 et 3.32.

diff. de F_0 moyen, en tons musicaux	diff. d'intensité moyenne, en dB	rapport : $\Delta I / \Delta F_0$
2.3	1.1	0.5 dB pour 1 ton musical

Tableau 3.32. Evaluation des différences entre tons lexicaux naxi M et H en termes de F_0 et d'intensité acoustique, locuteurs et conditions de lecture confondus.

Le tableau 3.32 indique qu'en naxi, **le rapport entre variation de F_0 et variation d'intensité est nettement plus élevé lorsque cette variation est due au ton lexical (différence entre un ton H et un ton M) que lorsqu'elle est due à l'intonation (différence entre condition de lecture soignée et condition de lecture insistante, à ton lexical constant)**. Ces observations suggèrent que la stratégie des locuteurs de langues à tons dans la mise en valeur locale d'une syllabe comporte un effort pour demeurer dans une certaine plage de fréquence fondamentale, celle du ton lexical de la syllabe.

Le ton L naxi n'est pas pris en compte dans ce calcul (qui représente une approximation assez grossière), du fait que le rapport entre F_0 et intensité n'est pas directement comparable entre le bas et le haut de la tessiture d'un même locuteur[1] ; pour le propos d'une mise en regard avec les syllabes accentuées anglaises, uniformément réalisées dans une plage de F_0 supérieure à celles du ton L du naxi, il paraît plus adéquat de ne retenir que la comparaison des tons H et M du naxi.

Les résultats qui ont été rapportés au cours de ce chapitre tiennent une place centrale dans la réflexion typologique menée dans la Discussion qui suit (chapitre IV).

[1] La différence entre ton M et L en naxi, conditions et locuteurs moyennés, est de 2 tons musicaux pour 4.1 dB.

Table des matières du chapitre IV (« Discussion générale »)

1. L'hypothèse d'un partage des ressources entre accentuation et intonation ____ 167

Mise au point de vocabulaire : intensification intonative et accent d'insistance ____ 169

1.1. La qualité de voix (mesurée par le quotient ouvert) varie moins en vietnamien, langue dans laquelle elle participe à une opposition lexicale ____ 170

1.2. Les rimes connaissent un allongement considérable en naxi, langue dans laquelle le trait de longueur n'est pas distinctif ____ 170

1.3. Les langues à tons contrôlent l'emploi intonatif de F_0 : le ratio entre variation de F_0 et variation d'intensité est plus bas que dans les données anglaises ____ 171

1.3.1. Le problème de la mise en valeur prosodique d'une syllabe portant un ton bas ou moyen, et une solution catégorielle ____ 171

1.3.2. La variation de F_0 lors d'un surcroît d'intensification intonative diffère d'une langue à l'autre ____ 173

1.3.3. Eléments de réflexion sur les mécanismes physiologiques sous-jacents à l'effort vocal ____ 173

1.3.4. Les résultats d'études sur d'autres langues paraissent confirmer l'hypothèse d'un partage des ressources entre accentuation lexicale et intonation ____ 175

2. Implications théoriques des observations : choix pour une modélisation prosodique ____ 178

2.1. Présentation de l'approche autosegmentale-métrique, et comparaison de l'emploi de la notion de ton dans les divers modèles ____ 178

2.1.1. Approche autosegmentale-métrique et traditions linguistiques américaines ____ 179

2.1.2. Influence des modèles de tonologie africaine sur l'approche autosegmentale-métrique ____ 180

2.1.3. Lien entre approche autosegmentale-métrique et traitement automatique des langues ____ 180

2.1.4. La notion de ton dans les modèles autosegmentaux-métriques et dans d'autres modèles (« britannique », « superpositionnel », « morphologique ») : un même mot, des emplois qui divergent ____ 180

2.1.5. Une version radicale du postulat des tons intonatifs : l'hypothèse d'une identité structurelle entre prosodie de langues à tons lexicaux et prosodie de langues sans tons lexicaux ____ 181

2.2. Le difficile débat entre approche autosegmentale-métrique, approche superpositionnelle, approche britannique et approche morphologique ____ 183

2.2.1. Principaux éléments du débat ____ 183

2.2.2. Résultats qui ne paraissent pas réductibles à une approche en termes de séquences tonales ____ 186

2.2.3. Réflexions sur l'idée d'une superposition ____ 188

2.2.4. Possibilités d'application didactique des modèles superpositionnels ____ 190

3. Réflexion typologique ____ 191

3.1. Quelques prototypes de systèmes prosodiques ____ 192

3.1.1. Accent libre et accent fixe ____ 192

3.1.2. Absence de phénomènes d'accentuation lexicale (« langues sans tons et sans accent de mot ») ____ 194

3.1.3. Systèmes de tons ponctuels ____ 194

3.1.3.1. Systèmes prosodiques africains à tons ponctuels ____ 195

3.1.3.2. Systèmes prosodiques asiatiques à tons ponctuels ; implications typologiques de leur description ____ 196

3.1.4. Registres de qualité de voix ____ 198

3.1.5. Tons modulés ____ 201

3.1.5.1. Le débat sur l'existence de tons modulés. Les propositions de réanalyse des tons modulés en tons ponctuels ____ 201

3.1.5.2. Apparition et évolution des tons modulés ____ 204

3.1.5.3. Un exemple développé de langue à tons modulés et qualité de voix : le vietnamien ____ 204

3.1.6. Entre système tonal et système accentuel : la question de l'accent dans les systèmes tonals, et les systèmes à accent mélodique (pitch accent) ____ 208

3.2. Une nouvelle perspective typologique, indépendante de la présence ou de l'absence de tons lexicaux : langues à *prosodie calculée* et langues à *prosodie contrôlée par l'énonciateur* ____ 214

Chapitre IV. Discussion générale

La discussion est divisée en trois volets, qui correspondent aux questions soulevées en introduction, reformulées ici sous forme d'hypothèses. Par ordre croissant d'abstraction :

Premier ensemble d'hypothèses, au plan des observations phonétiques : Les locuteurs de langues dont les tons sont définis par la hauteur recourent moins que les locuteurs de langues sans tons à F_0 dans la mise en valeur pragmatique de certaines syllabes ; en termes plus généraux, un *partage des ressources* a lieu entre spécification lexicale et emploi intonatif, les spécifications lexicales fournissant le cadre dans lequel la variation intonative prend son sens.

Deuxième ensemble d'hypothèses, au plan de la modélisation : Les mélodies de l'anglais (analysées en termes de tons dans les modèles autosegmentaux-métriques) diffèrent de façon considérable, au plan phonétique, des tons lexicaux du naxi et du vietnamien, ce qui amène à revenir sur certaines des prémisses des modèles autosegmentaux-métriques de l'intonation.

Troisième ensemble d'hypothèses, au plan typologique : le domaine des langues à tons comprend des systèmes prosodiques très variés, l'opposition entre langues avec et sans tons lexicaux ne permettant pas, par elle-même, de prédictions étendues sur l'intonation des langues concernées. La prise en compte d'un nombre relativement élevé de caractéristiques est nécessaire à l'établissement d'une typologie prosodique. En vue d'une contribution à la typologie ainsi envisagée, un paramètre nouveau pour la caractérisation des systèmes prosodiques est proposé.

1. L'hypothèse d'un partage des ressources entre accentuation et intonation

Les pages qui suivent partent des observations sur la réalisation phonétique de l'intensification intonative dans les trois langues étudiées, pour discuter la perspective d'un *partage des ressources* entre accentuation lexicale et intonation. Selon cette perspective, dans une langue donnée, **les moyens dont dispose l'énonciateur** (tels que : jeu sur la longueur, la hauteur, la qualité de voix) **jouent un rôle intonatif** *(post-lexical)* **inversement proportionnel à leur rôle de distinction lexicale** (Fry 1968 ; Vaissière 1983, 2004). Les locuteurs auront tendance à construire des stratégies intonatives qui n'interfèrent pas avec la bonne réalisation des spécifications lexicales. Par exemple, une langue où l'emploi de la longueur est contraint par la présence d'oppositions de longueur vocalique et de consonnes géminées (comme le finnois) emploiera moins la longueur au plan intonatif, tandis qu'une langue sans oppositions lexicales de longueur a toutes chances d'employer l'allongement dans le marquage du découpage en constituants « intono-syntaxiques » (cette dernière formule est reprise à Di Cristo 1985). Un point de vue similaire est exprimé par Klaus Kohler, qui, à l'occasion de tests de perception, relève l'emploi lexical, dans les oppositions tonales du chinois mandarin,

du contraste d'alignement des pics de F_0 qui en allemand (et dans d'autres langues à accent) distingue des catégories intonatives (expérience présentée et discutée dans l'Annexe 3, section 1) : lors de tests de perception, il apparaît que le déplacement du pic de F_0 réalisé sur la syllabe /lo/ dans la phrase *Sie hat ja gelogen* (« elle a menti »), déplacement qui pour une oreille allemande véhicule des nuances intonatives, donne aux auditeurs chinois l'impression d'un changement de ton. La différence est réinterprétée dans le cadre que fournissent les tons lexicaux du chinois mandarin (Kohler 1991c).

Ce raisonnement prédit que l'emploi intonatif de F_0 sera considérablement limité dans une langue à tons lexicaux ; que la qualité de voix sera utilisée au plan intonatif dans les langues dans lesquelles elle n'est pas distinctive, tandis que les langues à registre de voix (ou à tons combinant hauteur et registre de voix, comme le vietnamien), où la qualité de voix entre dans la définition des tons lexicaux, n'auraient pas la latitude de faire un usage intonatif de variations dans la qualité de voix.

Les expériences menées sur le naxi et le vietnamien établissent que les tons lexicaux, quel que soit leur nombre (quatre en naxi, six en vietnamien), laissent une forte marge de variation intonative. La présence de faits intonatifs variés dans les langues à tons d'Asie est connue depuis les travaux de Chao Yuen-ren 1933, 1935 sur le chinois[1]. Chao distingue trois niveaux : le ton lexical ; un niveau qui correspond à l'accentuation grammaticale (ce qui est désigné ici comme « intonation syntaxique ») ; et l'intonation *expressive*, qui inclut à la fois l'intonation pragmatique de ces auteurs, les attitudes et les émotions. Pour le seul domaine de l'intonation *expressive*, Chao Yuen-ren distingue pas moins de quarante types de réalisations. Le détail de son analyse appelle sans doute des retouches : en particulier, il pense que vingt-sept de ces types se distinguent exclusivement par la fréquence fondamentale, les treize autres engageant des différences d'intensité et de qualité de voix ; les recherches expérimentales réalisées depuis suggèrent qu'une distinction aussi tranchée est à nuancer. En revanche, la formulation de Kenneth Pike (qui a eu l'occasion de travailler avec Chao Yuen-ren) nous paraît garder toute son actualité :

> I would suggest that nonphonemic changes of pitch (…) constitute one of the most difficult characteristics of the Mandarin tonal system. Tonemes of syllables under emphatic stress tend to have wider spread of intervals : high pitches, there, are higher, low pitches are lower, falling glides begin higher and fall lower, and so on. The syllables other than the emphatically stressed one have the intervals between their pitches lessened : high tonemes tend to have their pitches lowered somewhat, low pitches are slightly raised, and glides become less long and prominent. In a sentence with an emphatically stressed syllable there may be considerable free

[1] L'orientation générale du travail de Chao Yuen-ren rappelle quelque peu la description du français par Pierre Delattre, qui distinguait des « intonations de base » : continuité mineure, continuité majeure, finalité, question, interrogation, ordre, exclamation, parenthèse haute, parenthèse basse, implication (Delattre 1966b).

> variation in the amount of modification of the unstressed tonemes : the glides, for example, are at times short, though still audible, or they become so abbreviated that I heard them only with close attention, or they are so drastically curtailed that I could detect no gliding pitch whatever. (Pike 1948:29)[1]

L'emploi lexical d'un paramètre n'est pas exclusif d'un jeu intonatif sur celui-ci. En particulier, l'emploi lexical de la fréquence fondamentale (dans la définition des tons lexicaux) n'est pas exclusif de ses emplois intonatifs ; F_0 a partie liée avec l'ensemble des phénomènes qui constituent l'intonation. **Néanmoins, les paramètres employés au plan intonatif varient quelque peu d'une langue à l'autre, cela d'une façon qui paraît corrélée avec la nature des oppositions lexicales présentes dans la langue.** Par exemple, F_0 varie significativement en naxi d'une condition de lecture à l'autre, mais le changement est d'une ampleur moindre qu'en anglais ; de même pour le quotient ouvert (indice de la qualité de voix) en vietnamien. **En d'autres termes, les phénomènes lexicaux (oppositions phonémiques, et accentuation lexicale) fournissent le cadre dans lequel la variation intonative prend son sens.** Ce constat est décliné ci-dessous pour les paramètres suivants : qualité de voix, longueur, formants, fréquence fondamentale et intensité.

Cet exposé est précédé d'une brève mise au point de vocabulaire.

Mise au point de vocabulaire : intensification intonative et accent d'insistance, marquage intonatif du focus, et focalisation contrastive

Sur la base des observations, et de ce qui nous paraît être un consensus émergent dans les travaux de pragmatique (en particulier : Paris 1998, 1999, Rialland et Robert 2001, Mazaudon 2003), il paraît utile de distinguer, d'une part, l'*intensification intonative*, possible à divers degrés dans toutes les langues, et dont l'accent d'insistance représente une forme extrême ; d'autre part le marquage intonatif du focus (autrement dit *morphème intonatif de focus*)[2]. Dans l'expérience principale, les mots-cible ont statut de focus dans l'une et l'autre condition de lecture ; une distinction entre intensification et marquage du focus paraît donc s'imposer pour décrire la différence entre conditions. Le marquage du focus est distinct de la *focalisation contrastive* : le naxi, à la différence de langues comme l'anglais, le vietnamien ou le chinois, ne connaît pas de focalisation contrastive. (Ce fait est établi par les tests de commutation employés pour délimiter le champ d'application du ton H de mise en valeur, chap. II, 2e volet, section 2.4.) En pratique, la distinction entre *marquage intonatif du focus*

---

[1] La variabilité des tons est telle que certains systèmes de reconnaissance automatique du chinois cantonais réalisent de meilleures performances lorsqu'ils n'utilisent pas l'information fournie par le module de reconnaissance des tons ; d'autres, par exemple le système proposé par Yao Qian, Tan Lee *et al.* 2004, parviennent à améliorer légèrement les performances grâce à la prise en compte des tons, mais l'amélioration demeure assez faible (de l'ordre de 8%).

[2] Pour plus de précisions, voir l'entrée « Focus » du Glossaire.

et *focalisation contrastive* n'est pas aisée à établir dans tous les cas ; néanmoins, c'est un fait typologique important que la focalisation contrastive est absente de beaucoup de langues.

1.1. La qualité de voix (mesurée par le quotient ouvert) varie moins en vietnamien, langue dans laquelle elle participe à une opposition lexicale

Les résultats des analyses statistiques menées sur le quotient ouvert en vietnamien permettent de considérer que **la qualité de voix varie significativement d'une condition de lecture à l'autre chez certains locuteurs, mais cette variation a lieu dans une plage beaucoup plus restreinte que chez les locuteurs naxi et anglais qui recourent à la qualité de voix dans leur stratégie d'insistance**. Apparaît en outre une remarquable similarité entre locuteurs : **en lecture insistante, les locuteurs mettent en valeur la qualité de voix lexicale du ton**, glottalisation au ton B2, voix plutôt détendue au ton D2, tandis qu'**en anglais et en naxi, où la qualité de voix n'est pas un trait phonémique, les stratégies individuelles qui recourent à ce paramètre sont opposées d'un locuteur à l'autre**.

1.2. Les rimes connaissent un allongement considérable en naxi, langue dans laquelle le trait de longueur n'est pas distinctif

La longueur de la rime varie significativement dans 50% des combinaisons locuteur-ton en vietnamien (exemple : ton D2 du locuteur D1 significativement plus court en lecture insistante), dans le sens d'un allongement chez certains locuteurs, dans le sens d'un abrègement dans d'autres ; en naxi, le chiffre est de 33% des combinaisons (toujours dans le sens d'un allongement) ; en anglais, la syllabe inaccentuée est plus brève en lecture insistante (chez 4 locuteurs sur 5), la syllabe accentuée a tendance à être allongée mais de façon moins nette (50% des combinaisons ; dans un cas, cette syllabe est abrégée en lecture insistante). Ce résultat a quelque chose de paradoxal : en naxi, où la longueur vocalique n'est pas un trait phonémique, une plus forte variation intonative de ce paramètre était attendue ; or les variations sont plus significatives en anglais et en vietnamien, où la longueur vocalique a rôle distinctif[1]. Néanmoins, ce résultat doit être relativisé au regard de

[1] Le trait phonémique de longueur n'a pas pour unique corrélat phonétique une différence de longueur : il s'accompagne d'une différence de timbre. Celle-ci est signalée par l'emploi de symboles différents pour les voyelles longues et les voyelles brèves dans les diverses notations de l'anglais en Alphabet Phonétique International : le *a long* de l'anglais est noté /ɑː/, son *a bref* /æ/ (voir par exemple Wells 1990, Viel et Lilly 1998b). Certains auteurs jugent que la différence entre voyelles longues et voyelles brèves appelle une description en termes de tension : un groupe de voyelles tendues s'opposerait à un groupe de voyelles relâchées, en fonction de la tension musculaire requise pour leur production : les voyelles de *beat* et *bait* seraient tendues, celles de *bit* et *bet* relâchées. Dans cette perspective, la différence de longueur n'a plus de réalité fonctionnelle (voir par exemple le manuel de Duchet et Fryd 1997, cité par Durand 2005). L'opposition tendu/relâché figure dans l'ouvrage désormais classique *Preliminaries to Speech Analysis* (Jakobson, Fant *et al.*

la longueur moyenne des syllabes (conditions confondues) : celle-ci est de 130 ms en anglais, 140 ms en vietnamien, et le double en naxi : 270 ms. En d'autres termes, dans les deux conditions de lecture, qui toutes deux donnent au mot-cible statut de focus, les locuteurs naxi réalisent une syllabe-cible très longue ; la différence relativement faible entre conditions provient du fait que l'allongement est fort dans l'une comme dans l'autre[1]. Ainsi, **le paramètre de longueur est employé abondamment au plan intonatif, comme le laissait prévoir l'absence de distinctions phonémiques de longueur en naxi**.

1.3. Les langues à tons contrôlent l'emploi intonatif de F_0 : le ratio entre variation de F_0 et variation d'intensité est plus bas que dans les données anglaises

1.3.1. Le problème de la mise en valeur prosodique d'une syllabe portant un ton bas ou moyen, et une solution catégorielle

Le problème de la bonne reconnaissance d'un ton lexical bas ou moyen lorsqu'il est modifié par des paramètres intonatifs qui le rehaussent paraît bel et bien se poser en pratique dans les langues à tons : la mise en valeur intonative des tons bas pose problème. En mandarin standard, langue qui est maintenant très abondamment décrite, le problème concerne le troisième ton, qui est réalisé dans une plage de fréquence fondamentale plus basse que celle des trois autres tons lexicaux (Hoa 1980:95, 237). Une observation similaire est faite par Shih Chilin et Kochanski 2000 : leur système de synthèse par ordinateur des tons du mandarin dans des énoncés complets repose sur des algorithmes qui déterminent la courbe de F_0 de la syllabe en fonction de son ton lexical et de son *degré de force*, ce dernier paramètre étant déterminé par l'intonation de la phrase ; ils constatent que la modification de la courbe du ton 3 selon les principes appliqués pour les autres tons aboutit en fait à un résultat jugé peu satisfaisant par les auditeurs, et qu'il est préférable (en l'état actuel des technologies de synthèse) de renoncer à appliquer ces manipulations à la courbe de F_0 du ton

---

1952), et son existence est aujourd'hui largement acceptée, dans une formulation qui a évolué vers une description en termes de Advanced Tongue Root ~ Retracted Tongue Root (par exemple Kenstowicz 1994:14). Néanmoins, cette analyse n'est pas universellement reçue : Jones 1918:39-40 émettait déjà des doutes sur l'utilité de l'analyse en termes de tension ; s'y ajoutent les arguments de Lass 1976, et tout récemment ceux de Jacques Durand, qui récapitule l'histoire de la notion (Durand 2005 « Tense/lax, the vowel system of English, and phonological theory »). La conclusion de Jacques Durand est la suivante : « ...phonologists should be more careful in their use of 'tense'-'lax'. (...) When one looks at the generalizations carefully, their only consistent interpretation is in terms of 'length' (whatever its precise theoretical translation) but somehow the traditional label tense/lax is maintained as denoting a further property of the opposition in question and furthermore taken as evidence for the universal relevance of the feature in question. »

[1] L'inspection visuelle du signal confirme que les syllabes-cible subissent un allongement en comparaison des syllabes de la phrase-cadre ; la mesure de la durée globale de l'énoncé permettrait une forme de normalisation, mais ce relevé n'a pas paru indispensable pour notre propos.

3. En naxi, le ton H possède une marge de liberté supérieure à celle des deux autres tons, du fait que le ton L est proche du bas de la tessiture du locuteur tandis que le ton H ne correspond pas au sommet de la tessiture : le locuteur est en mesure de réaliser une fréquence fondamentale nettement supérieure à celle que requiert la bonne réalisation d'un ton H. Rappelons que parmi les cinq niveaux intonatifs distingués par Rossi et Chafcouloff 1972 sur une base psychoacoustique, trois sont au-dessus du niveau *Médium : Suraigu, Aigu, Infra-aigu*, contre un seul au-dessous *(Grave)*. David Bradley (séminaire « Langues tibéto-birmanes », LACITO-Sorbonne Nouvelle) rapporte que dans une langue ni (yi), le ton H du déictique peut être fortement rehaussé à des fins intensives, avec pour effet le passage du sens de « là-bas » à « tout là-bas au loin », de « là-haut » à « tout là-haut là-haut ». En naxi de AS, le phénomène paraît lexicalisé sur un mot : /lo^{H}/[1] (« enjamber, franchir, traverser »), qui possède des emplois grammaticalisés (par exemple : /lo^{H} lɛ̄ bɯ̄/ « s'en retourner » ; /lɛ̄/ : « à nouveau » ; /bɯ̄/ : « aller »). Ce mot est nécessairement prononcé avec un contour très haut et fortement descendant, qui se distingue de celui des autres mots au ton H. En naxi de FK, ce sont trois adverbes qui présentent ce même ton *super-haut* lexicalisé : /de^{H}/, pour un objet lointain, sans précision de direction (/de^{H} ʈʂ^{h}ɯ̄ nɤ̀/ « là-bas au loin » ; /ʈʂ^{h}ɯ̄/ : déictique, /nɤ̀/ : « lieu, endroit ») ; /gɤ^{H}/, pour quelque chose d'éloigné en contrehaut ; /me^{H}/, pour quelque chose d'éloigné en contrebas. A l'opposé, aucun ton super-bas n'est apparu en naxi, ni (à notre connaissance) dans d'autres langues à tons, pour l'intensification d'un ton L. L'expérience menée au chapitre III confirme que l'intensification tend à rehausser tous les tons.

En naxi, le problème phonétique de la mise en valeur d'une syllabe au ton L peut être contourné, dans un certain nombre de cas, par le processus catégoriel d'ajout d'un ton H d'insistance (décrit au chapitre II, 2^{e} volet, section 2.4) : il porte précisément sur les syllabes dont le ton lexical est L ou M, catégories pour lesquelles une mise en valeur par rehaussement de plage de F_0 peut poser des problèmes d'intelligibilité. Le problème trouve là une solution catégorielle : la catégorie tonale de la syllabe est modifiée, par la création d'un contour montant, /LH/ ou /MH/ selon que la syllabe a un ton lexical L ou M. Néanmoins, ce processus ne représente que l'un des moyens offerts par la langue pour véhiculer des nuances de sens. Les sections suivantes reviennent sur les principaux résultats obtenus par l'expérience réalisée au chapitre III, qui concerne l'*intensification intonative*.

[1] Plutôt que de le noter avec un simple ton H, ou avec un double ton haut qui prêterait à confusion avec les tons flottants observés en naxi (évoqués au chapitre II, 2^{e} partie, section 2.2.7), le ton super-haut est ici noté par un H placé en exposant.

1.3.2. La variation de F_0 lors d'un surcroît d'intensification intonative diffère d'une langue à l'autre

Les résultats de l'expérience principale montrent que le paramètre de fréquence fondamentale varie de façon non catégorielle en naxi et en vietnamien aussi bien qu'en anglais ; en revanche, une différence apparaît entre l'anglais d'une part, le naxi et le vietnamien d'autre part : dans les langues à tons, l'augmentation d'intensité n'étant pas accompagnée d'une augmentation de F_0 de même ampleur qu'en anglais. Pour rappeler les principaux faits, la relation entre intensité et F_0 est du même ordre de grandeur en naxi et en vietnamien (de l'ordre de 3 dB pour un ton musical) ; elle est supérieure à ce qui est observé pour l'anglais (relation de l'ordre de 2 dB pour un ton musical), ce qui suggère que les locuteurs de langues à tons contrôlent dans une certaine mesure la fréquence fondamentale de telle sorte qu'elle tende à demeurer dans la plage du ton lexical de la syllabe. Cette observation rejoint une hypothèse qui a été formulée au sujet de langues d'Afrique à deux hauteurs tonales, dans lesquelles l'accent (grammatical) réalisé sur certaines syllabes n'affecterait pas la hauteur de la syllabe, qui serait déterminée quasi exclusivement par son ton lexical (Yves Moñino, séminaire LACITO, au sujet des langues gbaya-manza-ngbaka, famille niger-congo, groupe 6, adamawa-oubangui ; voir également Moñino 2001, « Le palenquero de Colombie: langue accentuelle ou langue tonale? »).

Soulignons que la variation intonative de F_0 dans les langues à tons étudiée est forte, et statistiquement significative dans la grande majorité des cas ; seule la comparaison entre langues fait ressortir les particularités de chacune. Il serait caricatural de dire que les locuteurs de langues à tons manipulent l'intensité indépendamment de F_0 : la relation demeure, mais de la masse des données émerge une tendance à une limitation relative de la variation intonative de F_0 dans les langues à tons.

*1.3.3. Eléments de réflexion sur les mécanismes physiologiques sous-jacents à l'*effort vocal

A la lumière de plusieurs publications sur les corrélats physiologiques de l'accent d'insistance, il paraît possible de formuler des hypothèses sur les mécanismes physiologiques sous-jacents aux variations que nous avons observées lors du passage de la condition de lecture soignée à la condition de lecture insistante. Pour rappeler dans ses très grandes lignes l'histoire des études sur ce que nous désignons comme intensification intonative, le point de vue le plus courant jusqu'au milieu du XX^{e} siècle est celui qu'exprime N.S. Trubetzkoy dans sa description de l'accent de phrase : « ... il s'agit toujours et seulement d'une gradation de la force expiratoire » (Trubetzkoy 1939 [1976]:242). Les résultats d'expériences de manipulation de l'intensité acoustique globale d'un signal enregistré ont conduit à prêter à l'intensité un rôle linguistique très secondaire en comparaison de F_0 (Fry 1955, 1958 ; Mol et Uhlenbeck 1956 ; Issatchenko et Schädlich 1966) ; ces résultats souvent cités (voir par exemple Bolinger 1965:17) ont amené à prêter une attention quasi exclusive à la fréquence fondamentale. Le paramètre d'intensité a été

quelque peu éclipsé des études prosodiques : pour ne prendre qu'un exemple, le terme n'apparaît pas dans le *Handbook of the Phonetic Sciences* (Hardcastle et Laver 1997), qui compte en revanche dix sous-sections pour l'entrée « fréquence fondamentale ». **Avec l'intensité, les questions de force articulatoire et d'effort vocal ont eu tendance à être reléguées au second plan** (bien qu'elles continuent par ailleurs à faire l'objet de recherches fondamentales : voir par exemple Traunmüller et Eriksson 2000 et références citées). De fait, il y a quelque imprudence à traiter de ces questions : les mesures acoustiques et électroglottographiques n'indiquent pas directement quels mécanismes physiologiques sont en jeu. La mise en regard des résultats avec ceux de plusieurs études sur les corrélats physiologiques de phénomènes semblables (et en particulier de l'accent d'insistance) suggère pourtant une hypothèse : **les divers travaux publiés sur la pression sous-glottique amènent à conjecturer qu'en lecture *insistante*, les sujets auraient tendance à accroître la pression sous-glottique, par une action accrue des muscles thoraciques (et ce dans les trois langues)**. Il est établi que, toutes choses égales par ailleurs, une augmentation de pression sous-glottique se traduit par une augmentation de F_0. On serait dans ce cas amené à supposer qu'en vietnamien comme en naxi, les locuteurs limiteraient l'augmentation de F_0 par un ajustement laryngien, pour qu'elle ne compromette pas l'identification du ton[1]. Cette piste de réflexion n'a pu être explorée plus avant du fait du caractère fortement invasif des techniques exploratoires dont l'apport serait décisif : mesure directe de la pression sous-glottique, et mesure de l'activité des muscles par électromyographie. Le quotient ouvert n'éclaire pas directement cette question : la différence entre conditions ne paraît pas varier significativement d'une langue à l'autre (par exemple, dans le sens d'une baisse de quotient ouvert en lecture insistante en naxi).

Dans le domaine plus large des études linguistiques, les études sur la notion d'*intensité* apportent des éléments de réflexion qui paraissent mériter l'attention des phonéticiens. Plusieurs études soulignent que **les sujets communiquants font flèche de tout bois dans l'expression de l'intensité** (voir par exemple Romero 2001, *L'intensité en français contemporain: analyse sémantique et pragmatique*). Ainsi, Hughes Constantin de Chanay avance « une définition *actancielle* de l'intensité comme *degré d'investissement de l'agent dans l'action et/ou de sollicitation des destinataires* » :

> ... la catégorie de l'intensité doit être envisagée comme une catégorie transversale, qui serait d'une certaine façon la même dans les composantes tant verbales que non-verbales [aussi dites *mimo-posturo-gestuelles*] de l'oral. (Constantin de Chanay 2005:229, « Pour une approche pluridimensionnelle de l'intensité »)

Dans le domaine spécifiquement phonétique, il a été mentionné (chapitre I, section 2.2.2) que des recherches récentes soulignent l'erreur de méthode des expériences de perception citées plus haut : l'intensité y était modifiée par un changement de gain global, qui affectait

[1] Des hypothèses à ce sujet avaient également été formulées au chapitre II, 2e partie, section 2.3.1.

également toutes les fréquences ; or dans la parole naturelle, l'augmentation de l'intensité vocale résulte de phénomènes plus complexes. « L'oreille opère une sommation des composantes spectrales différente de celle qui régit la mesure de l'intensité globale » (Rossi 1971:143) ; dans la parole naturelle, l'effort vocal s'accompagne, entre autre changements, d'une modification de la pente spectrale (paramètre dont l'importance est établie par de nombreux travaux : voir en particulier Sluijter, Van Heuven *et al.* 1997). **Il n'existe actuellement pas de méthode efficace pour la mesure de l'onde de débit glottique et de la pente spectrale ; en revanche, une hypothèse de travail raisonnable est qu'en parole naturelle, et chez un même locuteur, une augmentation notable d'intensité vocale (à partir de 3 dB environ) va de pair avec une augmentation d'amplitude relative des harmoniques supérieurs.** L'usage que nous faisons de l'intensité lui prête le statut d'*indice de l'effort vocal*, dans certaines conditions expérimentales s'entend : dans la comparaison des mêmes items sous diverses conditions de lecture.

1.3.4. Les résultats d'études sur d'autres langues paraissent confirmer l'hypothèse d'un partage des ressources entre accentuation lexicale et intonation

Les observations réalisées sur l'anglais, le naxi et le vietnamien appellent un élargissement à d'autres langues, en vue de vérifier si la variation respective de F_0 et de l'intensité acoustique globale suit des tendances différentes en fonction du système accentuel ou tonal de la langue concernée. Les données publiées sont nombreuses, mais toutes ne se prêtent pas aisément à une comparaison avec les nôtres[1]. Nous nous contenterons pour l'heure d'étendre brièvement la réflexion à une langue à accent mélodique, le suédois, et à une langue proche de l'anglais par son système accentuel : l'allemand.

L'idée générale selon laquelle l'intonation tend à faire usage des paramètres non spécifiés lexicalement prédit que les langues à accent mélodique, telles que le suédois, emploieront

---

[1] Ainsi, des différences dans la méthode expérimentale rendent difficile la mise en regard de nos résultats avec la comparaison de mesures de F_0, intensité et durée sur des syllabes accentuées dans sept langues proposée par Pamies Bertrán 1996, « Consideraciones sobre la marca acústica del acento fonológico ». Dans le domaine de la comparaison entre langues à tons et langues à accent, une étude récente de la parole lue en anglais et dans deux parlers chinois en usage à Taïwan (mandarin et min) compare également la variation de fréquence fondamentale et d'intensité dans ces trois langues, mais les mesures portent sur la plage de variation de fréquence fondamentale et d'intensité (Chen 2005, « The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects »). L'auteur conclut que les locuteurs de langues à tons recourent à une plage plus étendue de fréquence fondamentale et d'intensité, ce qu'il attribue à la présence des tons lexicaux ; l'étude montre également que la parole lue des locuteurs du parler min couvre une plage d'intensité acoustique plus étendue que celle des locuteurs du mandarin. Les prédictions que nous sommes amené à formuler sur la base des résultats obtenus ici ne portent pas sur la plage de variation de fréquence fondamentale et d'intensité, mais sur le degré de corrélation de ces deux paramètres, qui seraient relativement moins corrélés dans les langues à tons, cela dans le cas précis des syllabes qui sont porteuses d'un *morphème d'intonation pragmatique.*

moins F_0 à des fins intonatives que l'anglais[1]. **Les conclusions de certaines études paraissent en effet aller dans ce sens**, par exemple la comparaison entre l'anglais et le suédois menée par Eriksson, Grabe et Traunmüller (Eriksson, Grabe *et al.* 2002). Leur expérience consiste à faire écouter les mêmes stimuli (de langue suédoise) à des locuteurs du suédois et de l'anglais (ces derniers ne parlant pas suédois) ; les auditeurs doivent estimer le degré de saillance de chacune des syllabes de l'énoncé entendu. Sur la base de leurs réponses, les auteurs calculent un indice d'effort vocal : l'« effort vocal relatif apparent », agrégat d'intensité, de hauteur, et de la distance du locuteur au micro évaluée par un groupe d'auditeurs. Les auditeurs anglais attachent plus d'importance que les auditeurs suédois au paramètre de hauteur dans leur jugement sur le degré de saillance *(degré de mise en valeur pragmatique) ;* les auditeurs suédois, dont la langue maternelle possède des accents mélodiques, tiennent moins compte de la hauteur[2].

La prise en compte de l'allemand, langue typologiquement proche de l'anglais, peut aider à déterminer dans quelle mesure la variation relativement plus importante de F_0 en comparaison de l'intensité en anglais est une caractéristique propre à cette langue, et dans quelle mesure cette caractéristique est partagée par d'autres langues à accent de mot. Une étude récente (Gut 2000, « The phonetic production of emphasis by German learners of English ») conclut que les locuteurs allemands emploieraient plus l'intensité que les locuteurs anglais dans la réalisation de l'insistance *(emphasis)*. Notre conjecture est néanmoins que le groupe des langues à accent présente une certaine unité ; l'hypothèse formulée ici pourra être vérifiée par l'extension de l'expérience à d'autres langues à accent.

1.4. La variation des fréquences formantiques entre lecture soignée et lecture insistante est non significative dans la très grande majorité des cas

Erickson 2002 (« Articulation of extreme formant patterns for emphasized vowels ») établit que les fréquences formantiques varient significativement, en anglais, en fonction du degré d'insistance porté par la syllabe. L'étude de Donna Erickson porte sur des réalisations

[1] Rappelons que dans l'usage suivi ici, la prosodie se divise en *accentuation* et *intonation* ; aux yeux de lecteurs qui seraient accoutumés à employer *intonation* dans le sens que possède ici *prosodie*, recouvrant l'ensemble des faits « suprasegmentaux », accentuation comprise, la phrase qui précède deviendrait un curieux paradoxe.

[2] Cette expérience a bien sûr ses limites : les auditeurs suédois comprenaient les phrases, non les locuteurs anglais. Elle demanderait à être complétée d'une nouvelle expérience dans laquelle le sens des énoncés-stimuli ne seraient compréhensibles à aucun des deux groupes d'auditeurs. Ainsi, le fait que les auditeurs anglais recourent plus que les auditeurs suédois au paramètre de longueur vocalique dans leurs jugements pourrait s'expliquer de la façon suivante : les auditeurs suédois distinguent les voyelles longues des voyelles brèves, et ne sont pas tentés d'interpréter la différence phonétique de longueur entre ces deux ensembles comme un indice intonatif ; en revanche, les auditeurs anglais, incapables d'identifier les phonèmes vocaliques, auront tendance à attribuer la plus grande longueur de certaines voyelles à un allongement intonatif, participant à une stratégie de mise en valeur. (Voir les réflexions de Streefkerk 2002, *Prominence: acoustic and lexical/syntactic correlates*.)

fortement contrastées, à l'aide de paires d'énoncés telles que les deux réponses du tableau 4.1 :

Question	Answer
Put these two down?	No, put these two BACK.
Put these two back?	No, put THESE two back.

Tableau 4.1. Exemple de paires question-réponse employées par Donna Erickson dans une étude de la lecture insistante *(Erickson 2002).*

L'auteur emploie les termes de *unemphasized* vs. *emphasized* pour décrire les réalisations du /i:/ de *these* dans les deux réponses. L'opposition est ici entre un mot non accentué *(atonique)* d'une part, et le même mot porteur d'une focalisation contrastive d'autre part. Dans notre expérience, en revanche, aucune des deux conditions de lecture n'amène à une réalisation prosodique affaiblie de la syllabe : le mot-cible a statut de focus dans les deux conditions de lecture. La voyelle est donc réalisée pleine dans tous les cas (mise à part, bien évidemment, la voyelle /ə/ en syllabe non accentuée en anglais), ce qui distingue notre expérience de celle d'Erickson 2002 ainsi que de celles de Beckman et Edwards 1992 et de Jong 1995 (« The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation »). Cela explique que les valeurs de formants ne connaissent pas de changement notable d'une condition de lecture à l'autre. En anglais, les variations de valeurs formantiques sont moins systématiques que celles de hauteur, de quotient ouvert et d'intensité ; les cas sporadiques de variation significative observés ne se recoupent pas entre locuteurs. De même, en naxi, aucun schéma de variation commun à tous les locuteurs ne se dégage.

Les mesures de formants confirment par ailleurs la faible influence des tons sur le timbre des voyelles en naxi[1].

[1] Ces observations tendent à infirmer une hypothèse de R. Schulman (Schulman 1989, « Articulatory dynamics of loud and normal speech ») selon laquelle l'augmentation de la fréquence de F_1 constatée en « voix forte » serait une conséquence de l'augmentation de F_0, et répondrait à nécessité perceptive : la perception de la hauteur vocalique est influencée par la fréquence fondamentale, plus précisément par la distance entre F_1 et F_0 (Syrdal et Gopal 1986), d'où l'hypothèse selon laquelle une augmentation de fréquence fondamentale devrait s'accompagner d'une augmentation de la fréquence du premier formant pour que la qualité de la voyelle concernée soit préservée. (Cette hypothèse est notamment reprise dans une conférence d'Anna Geumann : Geumann 2001, « Vocal intensity: acoustic and articulatory correlates ».) Pour l'anglais, les prédictions de cette analyse sont compatibles avec les observations : F_0 augmente en condition de lecture insistante, F_1 augmente aussi quelque peu. Pour le naxi, cette analyse prédit une hiérarchie du F_1 selon le ton : le F_1 serait plus haut pour H que pour M que pour L. Les résultats ne sont pas conformes à cette prédiction, ce qui invite à revoir l'hypothèse sur laquelle elle est fondée : **une augmentation du premier formant concomitante avec une montée de F_0 n'est pas une nécessité universelle**, puisque les voyelles du naxi n'ont pas un premier formant plus élevé au ton H qu'au ton L. Il est imaginable qu'une compensation perceptive entre en jeu, de sorte que l'ajustement de F_1 au plan de la production selon le principe proposé par A. Geumann ne soit pas indispensable

2. Implications théoriques des observations : choix pour une modélisation prosodique

Le volet théorique de la discussion reprend et développe, à la lumière des résultats expérimentaux, ce qui a été dit en introduction au sujet du débat entre cadres d'analyse de l'intonation (en particulier l'approche *autosegmentale-métrique*, l'approche *britannique* et l'approche *superpositionnelle*). Il ressort que le point de vue extrême selon lequel la prosodie de toutes les langues pourrait se décrire intégralement à l'aide d'une séquence de tons (tons lexicaux et « tons intonatifs » se trouvant ainsi placés sur un même plan) soulève des difficultés de taille. Une réflexion est esquissée sur la conception de la prosodie comme superposition de plusieurs ordres de phénomènes.

2.1. Présentation de l'approche autosegmentale-métrique, et comparaison de l'emploi de la notion de ton dans les divers modèles

Le terme de « ton » est actuellement employé aussi bien pour décrire des phénomènes intonatifs (non lexicaux) que pour des phénomènes distinctifs au plan lexical. **Pierrehumbert et Liberman 1984:161 emploient l'expression programmatique de « English tonology » pour désigner l'étude de l'intonation anglaise**[1]. Un ouvrage collectif tout récent (*Prosodic typology: the phonology of intonation and phrasing*, dirigé par Sun-Ah Jun, Oxford University Press, 2005) présente des systèmes de notations en *tons* et *frontières* de l'intonation de onze langues[2], systèmes qui adoptent les principes de la notation ToBI *(Tones and Break Indices)* proposée par Silverman, Beckman *et al.* 1992 pour l'anglais, et codifié en détail par Beckman et Elam 1997[3]. **Les paragraphes qui suivent mettent en perspective ce cadre dominant[4], désigné ci-après comme autosegmental-métrique, le situant dans son**

à la bonne identification des voyelles. L'augmentation du premier formant qui ressort des données de cet auteur paraît être un des corrélats d'un effort vocal accru, non une conséquence de la hausse de F_0 en elle-même.

[1] L'idée est répandue dans les études autosegmentales-métriques ; pour ne prendre qu'un exemple, une communication de Brechtje Post (dont la thèse porte sur le français) est intitulée « French Tonal Structures » (Post 2002).

[2] Les langues en question sont l'anglais américain, l'allemand, le grec, l'italien, le serbo-croate, le japonais, le chinois mandarin, le coréen, le chinois cantonais, une langue amérindienne : le chickasaw, et une langue aborigène d'Australie : le bininj gun-wok (mayali). Le français ne figure pas dans cette liste.

[3] S'agissant de la composante BI, *Break Indices*, l'énoncé est découpé en deux niveaux de constituants : le groupe intonatif *(intonational phrase)* est lui-même divisé en groupes intermédiaires *(intermediate phrases) ;* un groupe intonatif se termine par un ton de frontière (noté H% ou L%), tandis que chaque groupe intermédiaire est marqué par un accent de groupe H- ou L- (qui peut être final ou initial). Ce système permet de transcrire un nombre élevé de configurations intonatives. Voir le site spécifiquement dédié à ToBI : http://www.ling.ohio-state.edu/~tobi/ame_tobi/. Nous ne débattons pas directement ici des vertus et des éventuelles limites de la transcription ToBI, la discussion portant essentiellement sur une réflexion de Robert Ladd présentée plus bas.

[4] Le statut dominant que possèdent actuellement les théories autosegmentales-métriques de l'intonation fait d'elles un *standard* de fait. Telle était bien la vocation de ToBI pour l'anglais (comme l'indique le titre de la

contexte d'apparition, et le comparant au courant britannique d'études intonatives, puis au courant *superpositionnel*.

2.1.1. Approche autosegmentale-métrique et traditions linguistiques américaines

Les études prosodiques qui décomposent l'intonation en séquences de tons (quatre niveaux tonals chez Leben 1976, « The tones in English intonation » ; deux tons chez Pierrehumbert 1980, *The Phonology and Phonetics of English Intonation*) sont dans la continuité des travaux de l'école américaine qui analysent l'intonation en plusieurs *niveaux* de hauteur, plutôt qu'en termes de *contours* (notamment Pike 1945, Harris 1951, Trager et Smith 1951 ; présentation synthétique par Lieberman 1967 ; Rossi 1999:16, 40 souligne la continuité de l'école américaine depuis Whitney 1867 et Bloomfield 1933). **Cette entreprise représente une tentative pour appliquer au domaine de l'intonation les méthodes qui ont fait leurs preuves dans le domaine des oppositions lexicales : déceler dans le domaine de l'intonation, qui paraissait soustrait à la *double articulation* qui caractérise le langage[1], un ordonnancement en catégories discrètes. Les unités descriptives employées sont des tons hauts (H) et bas (L : *Low*)** : les contours sont réanalysés en termes de séquences de tons ponctuels, « phonèmes de hauteur » selon la formule de Rossi 1999:27. Ainsi, la descente *(fall)* des descriptions britanniques est réanalysée comme une séquence HL.

---

communication fondatrice, « ToBI: A Standard for Labeling English Prosody », co-signée par Silverman, Beckman, Pitrelli, Ostendorf, Wightman, Price, Pierrehumbert et Hirschberg), puis pour d'autres langues. Ainsi, dans un article portant sur la prosodie du naxi, nous évoquions à titre de comparaison les phénomènes de *continuation* qui tiennent une place importante dans l'intonation du français, et citions Pierre Delattre ; l'un des relecteurs désignés par la revue *Diachronica* pour l'évaluation de cet article a critiqué la référence à Pierre Delattre comme trop ancienne, et suggéré de reformuler les faits dans le cadre autosegmental-métrique : « Why not reference more up-to-date A-M analyses? » La remarque nous a surpris, dans la mesure où la revue *Diachronica*, comme l'annonce son titre, est spécialisée dans les études diachroniques, plutôt que dans les discussions phonologiques (d'orientation synchronique) telles que celles qui entourent la modélisation de l'intonation. L'emploi de l'acronyme est révélateur : en présupposant la connaissance des théories en question, il revient à affirmer leur caractère incontournable.

[1] « On classe dans la prosodie tous les faits de parole qui n'entrent pas dans le cadre phonématique, c'est-à-dire ceux qui échappent, d'une façon ou d'une autre, à la deuxième articulation. Physiquement, il s'agit en général de faits phoniques nécessairement présents dans tout énoncé parlé : que l'énergie avec laquelle on articule soit considérable ou limitée, elle est toujours là, à un degré quelconque ; dès que la voix se fait entendre, il faut bien que les vibrations de la glotte aient une fréquence, ce qui donne à chaque instant, aussi longtemps que la voix est perçue, une hauteur mélodique déterminée ; un autre trait susceptible d'utilisation prosodique est la durée qui, bien entendu, est un aspect physique inéluctable de la parole puisque les énoncés se développent dans le temps. On comprendra, dans ces conditions, que linguistiquement ces faits ne puissent guère valoir par leur présence ou leur absence en un point, mais plutôt par leurs modalités, variables d'une partie à une autre d'un énoncé. En conséquence, ils se prêtent moins bien à caractériser des unités discrètes que d'autres, comme, par exemple, la nasalité ou l'occlusion labiale, qui peuvent figurer ou ne pas figurer dans un énoncé (...). » (Martinet 1960:83)

2.1.2. Influence des modèles de tonologie africaine sur l'approche autosegmentale-métrique

L'approche autosegmentale-métrique puise également une part importante de son inspiration (à commencer par le terme *autosegmental* lui-même) **dans les modèles issus de l'analyse de langues à tons lexicaux d'Afrique** (voir en particulier Rialland 1998, article qui retrace l'histoire de l'influence que l'étude des systèmes prosodiques africains a exercée sur le développement des théories phonologiques multilinéaires). Un apport de ces modèles est la séparation établie entre les tons et les voyelles avec lesquelles ils coïncident au plan phonétique : les tons ne sont pas attachés aux voyelles de la même façon que les traits qui définissent leur articulation (degré d'aperture, degré d'antériorité, et autres traits). Dans le cadre d'analyse autosegmental, il est proposé de représenter les tons comme des unités autonomes agencées sur une ligne *(tier)* particulière, parallèle à celle des unités de la ligne dite segmentale. L'analyse des tons vise à éclairer la façon dont ils s'associent aux unités de la ligne segmentale susceptibles de les porter *(tone-bearing units)*.

2.1.3. Lien entre le développement de l'approche autosegmentale-métrique et les besoins constatés dans le domaine du traitement automatique des langues

Le large succès de la notation ToBI est en partie lié au fait qu'elle paraissait coïncider avec les besoins constatés dans le domaine de l'étiquetage de corpus dans le traitement automatique des langues. Le caractère strictement linéaire de cette notation paraissait également propice à son emploi dans des systèmes de synthèse de courbes intonatives : les courbes sont stylisées, ne retenant que quelques points décisif *(cibles)* entre lesquels le détail de la courbe puisse être obtenu par une simple interpolation. Le choix de tons binaires H et L coïncide avec le mode de fonctionnement binaire des systèmes informatiques. Le paragraphe qui suit revient sur cette différence d'orientation entre modèle autosegmental-métrique et d'autres modèles prosodiques, en prenant pour exemple de ces derniers le modèle britannique.

2.1.4. La notion de ton dans les modèles autosegmentaux-métriques et dans d'autres modèles (« britannique », « superpositionnel », « morphologique ») : un même mot, des emplois qui divergent

Le terme de ton n'est pas nouveau dans les études intonatives : il était déjà employé de longue date, mais sans les implications qu'il possède dans le courant autosegmental-métrique. L'orientation bina riste de ce dernier courant signifie que **des notions apparemment semblables employées dans cette école et dans l'école britannique ou le courant superpositionnel s'inscrivent en réalité à l'intérieur d'architectures conceptuelles différentes, qui leur prêtent des significations éloignées**. **Les « tons » (*tones*, ou *tunes*) de l'école britannique sont des catégories descriptives définies sans *a priori* concernant leur nombre et les éventuelles sous-catégories existantes** : les plus courants sont la descente *(fall)*, la montée *(rise)*, la descente-remontée *(fall-rise)*, et la montée-redescente (*rise-fall*), chacun porteur de

connotations différentes ; des distinctions supplémentaires sont proposées entre *high-fall* et *low-fall*, *high-rise* et *low-rise*. La souplesse de ce cadres d'analyse autorise la reconnaissance d'autres configurations, moins courantes : *rise-fall-rise*, *level tone*. Des différences de détail existent entre les différents auteurs ; l'unité de cette école ressort néanmoins avec clarté par comparaison avec le courant autosegmental-métrique, courant guidé par l'attachement à une description en termes d'éléments discrets, programme dont l'aboutissement est la réduction de l'intonation à une séquence de tons H et L. De même, dans sa réflexion sur les modèles superpositionnels, Mario Rossi remarque :

> On rencontre en effet dans les travaux de l'Ecole d'Eindhoven les notions de tons Haut et Bas et de tons de frontière (Swerts 1997). Mais il s'agit en réalité d'abstractions pour désigner des points d'ancrage dans le continuum mélodique et donc d'unités abstraites du niveau de la substance, assimilables à des phonotypes. Une grammaire de l'intonation est constituée en réalité, dans le modèle IPO [Instituut voor Perceptie Onderzoek], d'un ensemble de mouvements mélodiques stylisés associés à des lignes de déclinaison de référence, et de règles combinatoires. (Rossi 1999:38)

Pour résumer, le terme de *ton* (tone), employé de longue date dans les études intonatives sur l'anglais (dans la tradition britannique et dans certains travaux superpositionnels) **n'implique pas d'allégeance à une analyse en tons binaires H et L**, pas plus qu'il n'implique l'affirmation d'une identité de nature entre les « tons » de l'intonation anglaise et les tons lexicaux de langues comme le yorùbá ou le chinois mandarin.

2.1.5. Une version radicale du postulat des tons intonatifs : l'hypothèse d'une identité structurelle entre prosodie de langues à tons lexicaux et prosodie de langues sans tons lexicaux

Yetunde Laniran évoque l'idée selon laquelle la prosodie des langues à tons lexicaux serait structurellement semblable à celle des langues à accent :

> It seems likely (...) that phonetic implementation would be based on similar principles regardless of a tone's origin. (...) There is no reason to believe that the rules for tone implementation have access to where the tones come from, since implementation follows the same general principles independent of language type (stress, accent or lexical tones). (Laniran 1992:4)

Robert Ladd développe la même idée, lui donnant son extension la plus générale. (C'est nous qui avons mis en gras certains passages.)

> In the most restrictive versions of current intonational phonology, it is explicitly assumed that independently chosen global shapes—e.g. a declination component—are not needed anywhere in the phonological description. **In effect, the restrictive linear view says that all languages have tonal strings; the main difference between languages with and without lexical tone is simply a matter of where the tonal**

> **specifications come from**. In some languages ("tone languages") the phonological form of morphemes often or always includes some tonal element, so that the tonal string in any given utterance is largely a consequence of the choice of lexical items.
> In this view, the only tonal elements that are free to serve pragmatic functions in a tone language are boundary tones, i.e. additional tonal specifications added on to the lexically determined tonal string at the edge of a phrase or utterance. (...)
> The principal *phonetic* difference between tone languages and intonation languages is simply a further consequence of the functional difference. More "happens" in F_0 contours in a tone language, because the tonal specifications occur nearly every syllable and the transitions span only milliseconds, whereas in a language like English most of the tonal specifications occur only on prominent words and the transitions may span several syllables. But **the specifications are the same kind of phonological entity regardless of their function, and transitions are the same kind of phonetic phenomenon irrespective of their length**. (Ladd 1992)

Ce point de vue, que Ladd, tout en le développant, prend soin de signaler comme *une des possibilités ouvertes* et non la seule (« In the most restrictive versions of current intonational phonology... »), ouvre une perspective vertigineuse. Il paraît accomplir, dans le domaine de l'intonation, le programme d'une linguistique qui soit une grammaire universelle, et *rende compte* du fonctionnement de toutes les langues. Cette conception est résumée par Di Cristo 2004:89 de la façon suivante : « Dans la mouvance de la phonologie autosegmentale, le ton est considéré comme une primitive de l'organisation phonologique de l'intonation. Il en découle que les langues à tons et les langues à intonation peuvent être décrites au moyen d'un même formalisme qui spécifie l'alignement des segments tonals avec le matériau lexical et supralexical. » (Au sujet de l'alignement de la courbe de fréquence fondamentale avec l'articulation, voir l'Annexe 3, section 1.) Le résumé de Di Cristo nous paraît faire ressortir avec netteté les confusions auxquelles donne lieu le formalisme qu'il évoque : l'opposition entre langues à tons et langues à intonation laisse à penser que les langues à tons sont des langues sans intonation, les séquences tonales étant déterminées par les tons lexicaux dans les premières, par l'intonation dans les secondes ; l'expérience de langues d'Asie comme le chinois ou le vietnamien révèle au contraire que ces langues possèdent, outre des tons, une intonation. L'opposition entre « langues à tons » et « langues à intonation » illustre de façon exemplaire le malentendu auquel peut conduire le point de vue exposé par Robert Ladd.

Il paraît intéressant de signaler que **deux des principales références du courant autosegmental-métrique, J. Pierrehumbert et M. Beckman, réservent en fait leur jugement sur ce point**. A propos d'une langue à tons, le hausa, elles écrivent :

> It is an open issue whether such languages have intonational systems that use the same representation and rule types as the more widely studied accent and stress languages. (Pierrehumbert et Beckman 1988:239)

Plus généralement, Janet Pierrehumbert remarque avec recul et avec une note d'humour que si les propositions contenues dans sa thèse de doctorat (*The Phonology and Phonetics of English Intonation,* Pierrehumbert 1980) pouvaient être appliquées telles quelles aux langues les plus variées, cela démontrerait son échec dans l'entreprise qu'elle s'était fixée : la description de l'intonation d'une langue particulière, l'anglais américain, et non d'une grammaire universelle de l'intonation. C'est sur cette question que porte la discussion qui suit : **l'approche expérimentale réalisée offre un moyen d'évaluer la validité de la thèse extrême formulée par R. Ladd, par la comparaison de données de deux langues à tons d'Asie avec des données anglaises, pour évaluer leur degré de similarité**. Au plan théorique, ce travail apporte des arguments dans le débat entre les conceptions autosegmentales-métriques, celles de l'école britannique, et les conceptions *superpositionnelles* (illustrées en particulier par les travaux de Grønnum 1991, 1992, 1998b, Gårding 1983, 1998, et en synthèse de la parole par Fujisaki et Nagashima 1969), pour qui l'intonation appelle une description en termes de superposition de plusieurs phénomènes.

2.2. Le difficile débat entre approche autosegmentale-métrique, approche superpositionnelle, approche britannique et approche morphologique

2.2.1. Principaux éléments du débat

Nina Grønnum formule dans les termes suivants le débat entre approche superpositionnelle et approche autosegmentale-métrique (qu'elle désigne comme approche « linéaire ») :

> The structure of intonation as ''superpositional'' *vs.* ''linear'' (or ''global'' *vs.* ''local'') has been debated for nearly two decades. The most pertinent aspects of the two types of intonation model can be summarized thus, ignoring differences of opinion within each major trend: a *superpositional* model describes intonation in terms of a layered system of simultaneous, interacting, non-categorial components of varying structural and temporal scope, where larger scope components carry and set the scale for smaller scope ones. The implementation of F_0 events is performed on the basis of upcoming as well as preceding events and is sensitive to syntactic and semantic structuring. A *linear* model describes intonation in terms of a sequence of categorially different, non-interacting tonal events (''pitch accents'' and ''edge tones'') whose manifestation is exclusively locally determined, implemented on a left-to-right basis, without look-ahead; there is no intonation component separate from the accents and the edge tones (...) (Grønnum 1998a:109).

Nina Grønnum souligne par ailleurs le **statut quelque peu incertain de la notion de « ton » telle qu'elle est employée dans le courant autosegmental-métrique d'études prosodiques : entité à la fois supérieurement abstraite et extrêmement concrète, fréquemment identifiée au paramètre phonétique qu'est le maximum local (pour les tons H) ou le minimum local (pour les tons L) de fréquence fondamentale.**

Le débat entre tenants et adversaires de la modélisation de l'intonation en tons ponctuels (nécessaire abstraction selon les uns, confusion rétrograde entre faits d'accentuation et faits d'intonation selon les autres) est rarement engagé de façon fructueuse. De part et d'autre existe le risque de caricaturer les thèses perçues comme adverses, ou tout simplement de ne pas en tenir compte. Bolinger tourne en dérision les partisans de la notation de l'intonation en niveaux (« Most American linguists, like American Rotarians and American baptists, are enthusiastic joiners » : Bolinger 1965:vii)[1]. L'échange de Nina Grønnum et Robert Ladd au sujet de l'ouvrage *Intonational Phonology* de ce dernier (Grønnum 1998a ; Ladd 1998) n'est guère plus constructif. Les tenants de l'approche autosegmentale-métrique reprochent à leurs adversaires de refuser de reconnaître des avancées objectives : Robert Ladd met en scène la victoire des modèles autosegmentaux, la présentation d'une thèse adverse étant suivie d'un « This view was convincingly challenged by... ». En retour, les adversaires de l'approche autosegmentale-métrique reprochent tout net à ce courant de perdre prise avec les faits, et de ne reconnaître dans les phénomènes que ce que postule leur modèle. Ainsi dans le commentaire de N. Grønnum au sujet de la description du hollandais par R. van den Berg, C. Gussenhoven et T. Rietveld :

> ...it will be apparent from what follows that I think van den Berg, Gussenhoven, and Rietveld's description obscures the most important fact about accentuation in Dutch, and that it tends to mispresent the relevant difference between contours in some instances because it disregards linguistic function (in a narrower as well as a wider sense). The purported phonological analysis thus nearly reduces to a phonetic transcription (though a broad one) and not always an adequate one at that, as far as I can judge. (Grønnum 1992:359)
>
> ...forcing Dutch intonation into a description in terms of sequences of categorically different, noninteracting pitch accents can be done only at the expense of phonetic (speaker/listener) reality » (*ibid.,* p. 367).

De même, Ivan Fónagy remarque que la perspective autosegmentale-métrique occulte la différence de nature qui sépare faits d'accentuation lexicale et faits intonatifs (point de vue qui rejoint celui de Martinet, l'intonation étant exclue de la double articulation qui caractérise les faits linguistiques au sens strict)[2].

---

[1] Voir notamment Martin 2001, « ToBI : l'illusion scientifique? », et d'autres communications au même colloque (Journées Prosodie 2001, Grenoble).

[2] Dans sa comparaison de GtoBI (« German ToBI ») avec le système de notation prosodique développé à l'Université de Kiel, Klaus Kohler note la difficulté qu'il y a à comparer les performances respectives de ces deux systèmes, du fait des prémisses théoriques différentes sur lesquels ils reposent, qui prêtent un sens différent aux catégories employées (Kohler 1995, « ToBIG and PROLAB: Two prosodic transcription systems for German compared » ; au sujet du système de notation prosodique de l'allemand PROLAB, voir Kohler 1992).

La chance typologique de la présente étude est qu'elle repose sur l'analyse d'une langue à tons ponctuels (le naxi) et d'une langue à tons complexes (le vietnamien), ce qui, espérons-nous, lèvera le soupçon de partialité en faveur de l'un ou l'autre camp. **Les quelques arguments mis en valeur dans le présent travail s'adressent spécifiquement à la thèse extrême évoquée par Robert Ladd, selon laquelle les tons ponctuels seraient les primitives de la description prosodique de toutes les langues.** Cette conception ne nous paraît pas adéquate pour la description de l'anglais britannique, non plus que pour celle des langues à tons d'Asie (des exemples précis sont fournis dans la section qui suit : section 2.2.2). Ce propos peut paraître radical ; il ne revient pas pour autant à remettre en cause l'intérêt des nombreux travaux réalisés dans le cadre autosegmental-métrique : ce courant fédère en réalité des chercheurs dont chacun apporte des innovations et des aménagements au modèle[1]. Barbara Kühnert (communication personnelle) compare l'entreprise de noter intonation et tons lexicaux par un même système à la tentative de caractériser consonnes et voyelles par les mêmes traits : cette dernière tentative offre un angle d'approche original pour prendre la mesure des similarités et des différences entre ces deux ensembles de phonèmes. Mariapaolo d'Imperio (exposé d'Habilitation à diriger des recherches, Sorbonne Nouvelle, 27 juin 2005), en réponse à Martine Grice qui lui demandait pourquoi elle ne jugeait pas opportun de remettre en cause les tons binaires H et L, qui soulèvent certaines difficultés descriptives, a souligné la nécessité d'un vocabulaire commun pour la bonne communication des idées entre chercheurs[2] ; l'argument est certes contestable, car il conduit à épouser ce qu'Albert Di Cristo désigne comme « consensus épidémiques », qui peuvent avoir pour effet d'entraver la réflexion ; en revanche, dans le détail de l'emploi qui est fait de la notion de ton binaire chez Mariapaola d'Imperio, il apparaît que des aménagements considérables sont introduits. En voici un exemple. En italien napolitain, la question serait caractérisée par une séquence L* + H, l'affirmation par L + H* ; la différence consisterait dans une différence d'alignement des mêmes cibles tonales (notion discutée dans l'Annexe 3, section 1). Mariapaola d'Imperio propose d'expliquer la différence d'allure constatée dans la courbe, non explicable directement par cette description, par l'idée que tons associés (H*, L*) et tons non associés (simples H et L) seraient séparés par une différence de nature : les tons associés connaîtraient une réalisation « forte », les tons non associés une réalisation « faible ». Un aménagement différent, mais dont les implications ne sont pas moindres, est

[1] Ainsi, Colin Wightman, co-auteur du texte fondateur du système ToBI, remarquait au congrès *Speech Prosody* : « in hindsight, the ToBI system contains several flaws that have limited its acceptance and application ». Le titre de sa communication (Wightman 2002) était « ToBI or not ToBI », bon mot qui circulait depuis quelque temps parmi les phonéticiens.

[2] De fait, les descriptions prosodiques qui ne s'inscrivaient pas dans le cadre dominant ont eu peu de retentissement ; pour ne prendre que deux exemples : les propositions déjà anciennes de Philippe Martin (Martin 1977b, 1977a) concernant la prise en compte de l'amplitude et de la durée des mouvements mélodiques, et récemment les propositions formulées par Jacqueline Vaissière, « Cross-linguistic prosodic transcription: French vs. English » (Vaissière 2002).

proposé par Laura Dilley, dans une thèse soutenue en 2004 : *The Phonetics and Phonology of Tonal Systems*. Voici une longue citation du résumé[1].

> This thesis examines the issue of descriptive adequacy in theories of the phonology and phonetics of intonation and tone. It is argued that in order to account for certain robust phonetic facts, such as the consistent presence and timing of F_0 peaks and valleys across languages (Ladd 1996), theories of the phonology and phonetics of tone and intonation must crucially specify the relative height relations of tones. For languages with lexically unconstrained placement of tones within the pitch range, such as English, an absence of restrictions on relative tone height is shown to lead to two types of problems: overgeneration of phonetic contours from phonological representations, and indeterminacy of phonological representations for phonetic pitch contours. It is shown that previous theories of intonation, including Pierrehumbert (1980), include insufficient constraints on relative tone height and therefore inherit these difficulties. To address these issues, it is proposed that relative height be represented in terms of structure known as a tone interval, which is an abstraction of a frequency ratio (e.g., a musical interval). (...)

La prise en compte des rapports de hauteur ouvre incontestablement des perspectives intéressantes. Il paraît certes dommage que la discussion ainsi ouverte ne sorte pas du face-à-face avec les mêmes chercheurs (Janet Pierrehumbert et Robert Ladd), et ne remette pas en cause leur prémisse selon laquelle, à un niveau sous-jacent, l'intonation consiste en une séquence de tons binaires H et L : une des conséquences de cette orientation est que le volet expérimental de la thèse perpétue l'équation entre intonation et F_0, plutôt que de faire une place à la variété des mécanismes physiologiques qui entrent en jeu. Néanmoins, les propositions avancées par L. Dilley illustrent la **complexité des travaux qui s'inscrivent dans le courant autosegmental-métrique, dans lesquels le postulat de tons H et L tient parfois une place moins essentielle qu'il n'y paraît de prime abord**.

2.2.2. Résultats expérimentaux qui ne paraissent pas réductibles à une approche en termes de séquences tonales

Ce paragraphe met en perspective certains résultats de la présente étude qui nous paraissent plaider pour une approche en termes d'interaction de plusieurs ordres de phénomènes plutôt qu'en termes de séquences de tons.

2.2.2.1. Observations sur le marquage des frontières en naxi

En naxi, les frontières sont marquées par un allongement, et fréquemment par une descente. L'étude phonétique détaillée de la réalisation des tons lexicaux et des frontières suggère que ces dernières ne se traduisent pas par l'ajout, dans la chaîne des tons, de tons H, M ou L

[1] Le texte complet de la thèse est actuellement disponible à l'adresse suivante : http://stuff.mit.edu/afs/athena.mit.edu/user/d/i/dilley/www/publications.htm

supplémentaires. Cette observation s'inscrit en faux contre l'hypothèse que Robert Ladd baptise « restrictive linear view », selon laquelle tons lexicaux et frontières contribueraient en définitive à composer une unique séquence tonale pour l'énoncé. A un degré plus élevé d'abstraction, il demeure bien sûr possible de parler d'un « ton de frontière », précisant que sa réalisation se distingue de celle d'un ton lexical ; mais il paraît plus avisé de parler simplement de *frontière*, pour refléter la différence de nature constatée entre tons et frontières.

2.2.2.2. Divergence entre nos observations et celles de Pierrehumbert et Liberman sur l'insistance en anglais

Janet Pierrehumbert et Philip Liberman avancent l'idée selon laquelle l'insistance *(emphasis)* consiste en un changement scalaire appliqué à des « tons intonatifs » présumés catégoriels (*High* suivi de *Low* ; voir Pierrehumbert et Liberman 1984). Ils constatent une cible haute de plus en plus élevée et une cible basse constante, donc une descente accrue entre point le plus haut et point le plus bas, à mesure que croît le degré d'insistance. Francis Nolan, reprenant de façon assez fidèle le protocole expérimental de Pierrehumbert et Liberman, souligne les difficultés que soulève le caractère strictement linéaire de leur modèle, qu'il désigne comme TS, pour « Tone Sequence », et auquel il oppose une approche qu'il nomme CI, pour « Contour Interaction » (Nolan 1995 suit en cela la terminologie proposée par Ladd 1983a ; il associe l'approche « Contour Interaction » avec les travaux de l'école britannique : Palmer 1922, Kingdon 1958, Crystal 1969, O'Connor et Arnold 1973). **Nos résultats suggèrent, plus généralement, que la modélisation de l'intonation en tons soulève des difficultés dans les trois langues étudiées.** Notre protocole expérimental n'est certes pas identique à celui mis en place par Pierrehumbert et Liberman 1984, qui imposent à leurs quatre locuteurs de distinguer dix paliers à l'intérieur d'un continuum de "overall emphasis or excitement" (il n'est pas fourni d'indications de contexte ; cette expérience est évoquée plus en détail en tête du chapitre III). Néanmoins, **l'emploi d'un protocole expérimental similaire pour les trois langues étudiées ici permet une comparaison, laquelle fait ressortir le degré de variation des courbes mélodiques chez les sujets anglais**. L'absence de différence systématique de pente entre conditions de lecture sur les monosyllabes anglais ne paraît pas compatible avec l'hypothèse selon laquelle ils seraient porteurs d'un contour /HL/, comme le proposent Pierrehumbert et Liberman 1984, selon qui le ton L serait réalisé à une hauteur constante tandis que le ton H serait rehaussé proportionnellement au degré d'insistance. La variabilité dans les contours de fréquence fondamentale observés ne nous semble pas pouvoir être ramenée à une telle modification scalaire de tons catégoriels. Poursuivant l'argument plus avant, il nous paraît probable qu'une recherche inductive qui partirait de nos données anglaises pour y déceler des configurations tonales ne parviendrait pas à saisir l'unité du phénomène prosodique qui se réalise chez les divers locuteurs.

2.2.2.3. Bilan

L'« Epilogue » de l'article de synthèse de Di Cristo 2004 contient la réflexion suivante :

> Si l'investigation des corrélats physiques et leur traitement statistique a pu faire l'objet d'analyses de plus en plus abouties, les propositions se rapportant à la nature des *représentations mentales des systèmes prosodiques* demeurent hautement spéculatives et fortement dépendantes des présupposés théoriques. Nous devons convenir que ces représentations ne sont que des « représentations de représentations de représentations » (Culioli 1990 ; Auran 2004), dans la mesure où elles représentent ce que les linguistes se représentent être les représentations mentales du système cognitif. Il n'est pas interdit de penser que les systèmes de représentation qui s'imposent aujourd'hui comme les plus influents et qui suscitent, de ce fait, les plus larges consensus épidémiques, se verront supplantés dans un avenir plus ou moins proche (...) (jusqu'à quel point peut-on affirmer aujourd'hui avec sérénité que l'intonation est représentée mentalement par des suites de segments tonals ?) (p. 184)

Au vu des observations réalisées, notre réponse à la question que soulève (avec diplomatie) Albert Di Cristo à la fin de ce passage serait qu'**il est préférable de faire l'économie de la notion de ton ponctuel binaire dans l'analyse de l'intonation**. Avec plus d'humour et de nuance, Jacqueline Vaissière [séminaire « Jeux prosodiques »] se déclare partisan de ToBI, pour peu que l'acronyme soit *accentué à la française* : non pas avec l'accent initial que lui donne la langue anglaise (/ˈtəʊbɪ/), accent qui tombe sur la composante *Tones*, mais avec une accentuation finale ([toˈbiː] !), à la française, qui porterait sur la partie BI, *Break-Indices*, celle des frontières ; façon de dire que les « tons intonatifs » gagneraient à passer à l'arrière-plan des descriptions. **Le terme de *ton*, qui dans les travaux de l'école autosegmentale-métrique est à la fois une primitive de la description phonologique et un élément employé dans la stylisation des courbes de F_0** (comme les signes de l'Alphabet Phonétique International dans une notation sub-phonémique), **appellerait remplacement par *morphème intonatif*** terme qui a le double avantage d'éviter la confusion avec les tons lexicaux et de représenter clairement une catégorie de forme, sans préjuger des réalisations phonétiques (Rossi 1999). En l'état actuel des débats, qui demeurent dominés par le modèle autosegmental-métrique, le propos risque d'être jugé rétrograde ; il n'est pas inimaginable que l'évolution des modèles autosegmentaux-métriques les conduise néanmoins dans cette direction.

2.2.3. Réflexions sur l'idée d'une superposition

Dans le domaine des langues à tons d'Asie, le linguiste est conduit à distinguer entre les tons et les divers phénomènes intonatifs ; en témoignent les travaux d'Abramson 1978, 1979 ; Gsell 1979b ; Luksaneeyanawin 1983 sur le thai, et de Chao Yuen-ren 1933 ; Hoa 1980 ; Shih Chilin 2000 sur le mandarin. **Un terme revient fréquemment dans les descriptions :**

celui de *superposition*. Une citation tirée d'une description sommaire de l'intonation du dialecte chinois du Sichuan résume l'idée générale :

> Intonation does exist in the Chengtu dialect. It is superimposed on the sentence as a whole. And it is this superimposed intonation that modifies the individual tones, and not the tones that decide the intonation of the sentence. (Chang Nien-Chuang 1958:84)

De même chez Monique Hoa :

> En partant des tons comme base, on peut dire que l'intonation se superpose aux tons pour donner le produit final. (Hoa 1980:8)

L'idée d'une superposition de plusieurs ordres de phénomènes prosodiques est également répandue parmi les chercheurs qui ont travaillé à la synthèse par règles de la prosodie, en particulier Fujisaki et Nagashima 1969, Vaissière 1971, Lea 1973, Carlson et Granström 1973. La notion de *superposition* est devenue éponyme d'un courant de recherche, associé aux noms de Nina Grønnum (qui a également publié sous le nom de Thorsen[1] ; cf Thorsen 1984, Grønnum 1991, 1992, 1998a) et Eva Gårding (Gårding 1983, 1998)[2]. Dans cette perspective, les événements prosodiques locaux, et les phénomènes *microprosodiques* (influence des segments sur la mélodie), se superposent à une courbe globale définie au niveau de l'énoncé tout entier, et de chaque groupe intonatif.

> ... I have suggested a description of the F_0 pattern characterizing prosodic stress groups in Danish in terms of an essentially invariant F_0 wave upon which the segments and syllables are superposed. This will make intrasyllabic F_0 movements predictable from the shape of the wave where the syllables hit it (falling, rising, etc.). Any F_0 movement may then of course be modified by microprosodic (segmental) phenomena. (Thorsen 1984:17)

L'idée d'une superposition n'est pas exclusive de celle d'une interaction entre les composantes superposées : ainsi, les données montrent qu'un même degré d'insistance se traduit différemment en fonction du ton lexical d'une syllabe naxi ou vietnamienne ; autrement dit, la réalisation du morphème intonatif d'insistance n'est pas aveugle à l'identité du ton lexical auquel ce morphème se superpose. **Les modèles superpositionnels soulignent les similitudes entre les modalités d'opération de l'intonation dans diverses langues, en même temps qu'ils font ressortir les implications qu'a, pour l'ensemble de la prosodie, la présence de certains faits d'accentuation lexicale (tels que les tons lexicaux)**.

Le dialogue ouvert avec les idées superpositionnels ne représente nullement une allégeance exclusive à tel ou tel modèle en particulier ; les idées superpositionnelles nous paraissent

---

[1] Sur sa page personnelle, cet auteur précise : « Until the mid 1980s I published under my married name, Thorsen, and for a couple of years also as Grønnum Thorsen » (http://www.cphling.dk/~ng/publ-90.html).

[2] Un autre travail superpositionnel fondateur est celui d'Öhman 1967.

présenter des similarités essentielles avec celles d'autres traditions, dont la tradition d'études intonatives britanniques, et ce que Rossi nomme « théories morphologiques de l'intonation » (Rossi 1999). L'intitulé circonspect de la présente section (« Réflexions sur l'idée d'une superposition ») tient au fait que nos expériences sont limitées à l'analyse de données de production (et secondairement de perception), et n'ont pas débouché sur une application en synthèse de courbes de F_0, étape décisive dans la modélisation[1]. Nous ne sommes donc en mesure de proposer qu'une réflexion sur les possibilités ouvertes pour la modélisation. En revanche, l'intuition selon laquelle les modèles développés pour les langues subsahariennes (puis exploités dans la description de l'intonation de l'anglais et d'autres langues) ne conviennent pas à la description des langues à tons d'Asie que sont le naxi et le vietnamien ouvre des perspectives typologiques. **Les trois langues étudiées ont ceci de commun que le choix d'une réalisation phonétique plus ou moins saillante des phénomènes accentuels** (accent en anglais, tons en naxi et vietnamien) **est à tout instant sous le contrôle de l'énonciateur**, ce qui n'est pas le cas dans de nombreuses langues subsahariennes, pour lesquelles nous proposons ci-dessous (dans la section 3, consacrée à une réflexion typologique) la dénomination *langues à prosodie calculée*.

2.2.4. Possibilités d'application didactique des modèles superpositionnels

Le présent travail se plaçait d'emblée d'un point de vue de recherche fondamentale, plutôt que d'application à l'enseignement. Recherche et enseignement n'en permettent en réalité un *éclairage croisé* profitable à l'un et l'autre domaine ; les quelques remarques ci-dessous visent à ouvrir quelques perspectives.

S'il était permis, dans ce travail de recherche fondamentale, de tirer argument d'applications didactiques, nous aimerions souligner que **le cadre superpositionnel correspond à des méthodes courantes d'enseignement de la prosodie**. Dans son apprentissage de la prononciation d'une langue étrangère, l'apprenant est invité à travailler séparément les phonèmes (à l'aide de paires minimales et de monosyllabes transcrits en alphabet phonétique international), puis l'accentuation du mot (par la lecture de mots isolés), et enfin la composante syntaxique de l'intonation et sa composante pragmatique (par la lecture de phrases). Dans l'enseignement de l'anglais, par exemple, manuels et années d'étude (et jusqu'à certaines épreuves d'examen) suivent cette division, le travail sur l'intonation

[1] La synthèse de la parole permet la vérification d'hypothèses formulées sur la base d'observations de parole naturelle, cela depuis le *Pattern Playback*, qui, dès les années 1950, a permis l'établissement de l'importance des transitions formantiques pour la perception des consonnes (Delattre 1958, 1965, 1966b). La simulation réaliste de la source de voisement demeure difficile : le contrôle de l'*onde de débit glottique*, du *quotient ouvert* et du *quotient de rapidité* soulève toujours des difficultés de taille (sur ces notions, voir Henrich 2001), bien que les avancées les plus récentes permettent de jouer sur certains paramètres glottiques de façon indépendante, permettant par exemple à Abramson, Thongkum *et al.* 2004 des expériences sur la perception des registres de voix (voix modale, voix soufflée/*breathy voice*).

succédant au travail sur l'accentuation. (Voir entre autres exemples Viel et Lilly 1998a, *Initiation raisonnée à la phonétique de l'anglais*.) En cela, la pratique didactique est éminemment compatible avec le cadre superpositionnel, et les idées superpositionnelles paraissent se prêter sans difficulté aucune à une application didactique. Le fait que l'enseignement soit en un sens en avance sur la théorie représente autant un encouragement dans la recherche qu'un **gage de succès dans l'application de celle-ci**.

S'agissant de l'apport de nos données expérimentales à la réflexion sur l'enseignement de la prosodie de l'anglais, elles plaident pour une description dans laquelle les événements prosodiques sont présentés comme rattachés à des unités plus larges que la syllabe. En cela, il nous paraît profitable de prendre pour base, dans l'enseignement, la conception *britannique* des « tons » (« tones » ou « tunes ») : comme il a été souligné plus haut (section 2.1.4), le terme de ton *(tone)*, employé de longue date dans les études intonatives sur l'anglais, en l'absence de tons lexicaux dans cette langue, n'est, dans la tradition britannique, qu'une étiquette commode : il peut être remplacé par celui de *contour*, ou de *tune*. **L'emploi du terme de *ton* (dans cette approche et dans certains travaux superpositionnels) n'implique pas d'allégeance à une analyse en tons binaires H et L rattachés à telle ou telle syllabe, pas plus qu'il n'implique l'affirmation d'une identité de nature entre les « tons » de l'intonation anglaise et les tons lexicaux de langues comme le yorùbá ou le chinois mandarin**[1].

3. Réflexion typologique

Il y a quelque imprudence à s'engager dans un tableau typologique. Les descriptions de langues sont parfois difficiles à rapporter les unes aux autres, du fait de différences dans le vocabulaire employé dans les travaux des spécialistes des langues concernées.

> Over the past several decades, many linguists have made pronouncements on how prosodic systems should be typologized. While no other area of phonological typology has attracted as much attention, it is not clear how much consensus there is on where things currently stand. Linguists sometimes speak or write on the subject with great conviction, but the views that are expressed often conflict on both general and specific issues, e.g. whether all human languages must have "stress", whether the prosodic system of Tokyo Japanese is "tonal" vs. "accentual", and so forth. (Hyman 2005:1)

[1] Ainsi, dans l'école britannique, le fait que certaines mélodies s'étendent au-delà du noyau *(nucleus)* et même au-delà du mot accentué ne soulève pas de difficulté théorique. O'Connor et Arnold signalent notamment : « In the Low Rise and High Rise the stressed syllable of the accented word does not itself rise in pitch, but each of the following syllables is a step higher than the previous one. (…) When the final syllable of the word group is stressed (and only then), there is a tendency to slide upwards in pitch during the syllable » (O'Connor et Arnold 1973:12-13).

Cet exercice délicat nous paraît néanmoins essentiel pour que la recherche linguistique réalise des progrès cumulatifs.

3.1. Quelques prototypes de systèmes prosodiques

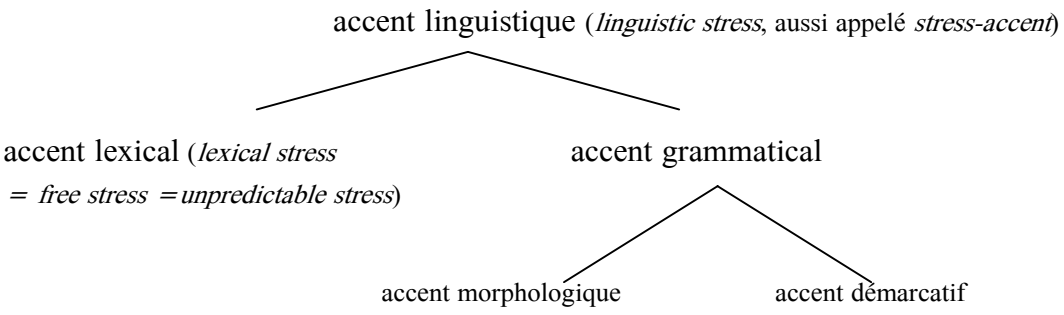
Un examen typologique révèle la diversité interne des grandes catégories que sont les systèmes accentuels, les systèmes de tons ponctuels, de registre de qualité de voix, de tons modulés, et d'accents mélodiques. Il n'en demeure pas moins utile d'entamer la réflexion typologique par un panorama organisé par prototypes de systèmes prosodiques[1] ; tel est l'objet de la présente section (3.1). Une nouvelle perspective typologique sera ensuite ouverte (section 3.2).

3.1.1. Accent libre et accent fixe

Dans une perspective typologique, il paraît important de maintenir une distinction entre accentuation lexicale (libre) et accentuation non lexicale, comme le soulignait déjà Trubetzkoy :

> *Corrélation d'accentuation* Dans ce chapitre, consacré aux fonctions phoniques distinctives, il ne peut naturellement être question que de l'accentuation dite « libre », c'est-à-dire seulement d'une accentuation dont la place dans le mot n'est pas conditionnée extérieurement et qui éventuellement peut différencier des significations du mot (par exemple russe mùkă « tourment, supplice » — mŭkà « farine »). (...) Ce qui est phonologiquement essentiel pour les langues à accentuation libre, c'est premièrement que cette mise en relief n'a lieu dans chaque mot qu'à une seule place, (...) et deuxièmement que, dans les mots ayant le même nombre de prosodèmes, la mise en relief n'affecte pas toujours le même prosodème, de sorte qu'il peut exister des paires de mots qui se distinguent l'un de l'autre uniquement par la place du sommet accentuel. (Trubetzkoy 1939 [1976]:221)

Les catégories proposées par Hyman 1977, et retenues dans la présente discussion, sont les suivantes :



[1] Comme le remarque notamment Anthony Fox, la typologie ne saurait être que prosodique et non intonative : une typologie purement intonative, qui tenterait de faire abstraction des faits d'accentuation, ne paraîtrait guère prometteuse (Fox 1995, « Principles of intonational typology »).

> In any given language stress may function in any one or more of these, and in fact, the typical case is that a language with lexical stress may also have some grammatically predictable stress, and that a language with demarcative stress may be occasionally upset by morphological considerations. The final stress in *terminó*, for instance, can be grammatically predicted on the basis of the *–ó* suffix meaning 'third person singular past tense', which always takes stress in Spanish. Other stress oppositions are lexically defined in Spanish (e.g. *término* 'term'), while the general (demarcative) pattern is penultimate stress (e.g. *termíno*, 'I terminate'). (Hyman 1977:39)

Ces remarques de Larry Hyman rappellent que **l'opposition entre langues avec et sans accent libre n'est pas absolue**. L'accent fixe du latin se plaçait sur la pénultième, sauf si celle-ci était brève, auquel cas il se plaçait sur l'antépénultième ; du latin à l'italien moderne, la perte des oppositions de longueur vocalique a rendu non prévisible la position de l'accent, qui de ce fait a acquis statut distinctif au plan lexical. « Many cases of the development of lexical from grammatical stress have been reported in the literature » (Hyman 1977:40)[1].

Le vaste ensemble des systèmes prosodiques accentuels (« langues à accent ») présente une assez grande variabilité interne[2] (voir par exemple les travaux expérimentaux d'Anne Cutler, et le projet Stress Typology[3], qui offre une table d'orientation des très nombreux travaux concernant la prosodie des langues à accent). Le rendement fonctionnel de l'accent varie fortement d'une langue à l'autre, selon que l'accent y est plus ou moins prédictible (voir, pour l'anglais, Chomsky et Halle 1968) ; plus il est prédictible, plus il tend à jouer un rôle simplement démarcatif.

---

[1] Pour le français comme pour l'italien, Mario Rossi maintient l'emploi du terme d'*accent lexical*, dans les termes suivants : « L'accent lexical est une propriété virtuelle du morphème lexical, dont le domaine est la syllabe et dont l'organisation syntagmatique crée un cadre intégrateur du mot » (Rossi 1999:117-118, qui renvoie à Rossi 1967, 1977, 1979, 1981).

[2] Ainsi, l'espagnol, langue à accent, serait plus proche du français que des langues germaniques par son rythme : « The values obtained agree with the classification of English, Dutch and German as stress-timed and French and Spanish as syllable-timed » (Grabe et Low 2002:515) ; ces classifications ont néanmoins leurs limites : voir les remarques générales de Dauer 1983, « Stress timing and syllable timing reanalysed » ; au sujet de l'espagnol : « Spanish is characterized as neither stress-timed nor syllable-timed, as the chief operative unit is the tonic group, which unites a stressed syllable with preceding unstressed syllables » (Alcoba et Murillo 1998) ; voir également Kvavik et Olsen 1974, « Theories and methods in Spanish intonation studies », et Pointon 1980, « Is Spanish really syllable-timed? ». Tomáš Duběda—exposé à l'Université Paris 8, et livre à paraître—signale que les méthodes de Franck Ramus (Ramus, Dupoux *et al.* 2003) et de Grabe et Low *(op. cit.)* aboutissent à des résultats opposés pour la langue tchèque.

[3] http://www.let.leidenuniv.nl/ulcl/pil/stresstyp

3.1.2. Absence de phénomènes d'accentuation lexicale (« langues sans tons et sans accent de mot »)

Dans les synthèses typologiques, le français est assez souvent décrit comme langue à accent final (voir par exemple Hyman 1977:64). En réalité, le français, langue sans accent lexical distinctif, ne possède pas clairement d'accent démarcatif de mot, la démarcation ayant lieu de façon plus saillante en fin de *groupes intonatifs* qu'en fin de mots (Delattre 1966a). Cette langue fait figure d'exception parmi les langues nationales européennes (romanes, germaniques ou slaves). Entre autres études, Rossi 1979 posait à nouveau la question : « Le français, langue sans accent ? » Jacqueline Vaissière propose l'appellation de « langue à frontière » pour le français (Vaissière 2002, « Cross-linguistic prosodic transcription: French vs. English ») ; la catégorie a ceci d'insatisfaisant que toutes les langues connues possèdent pareillement des frontières ; de la même façon qu'opposer « langues à tons » et « langues à intonation » tend à perpétuer l'illusion selon laquelle les langues à tons lexicaux n'auraient pas d'intonation. **Une définition négative paraît offrir la caractérisation la plus claire de ce type prosodique : *système prosodique sans tons, sans accent lexical distinctif, et sans accent démarcatif de mot*.** Le français n'est pas un cas isolé : les langues répondant à ce signalement prosodique seraient notamment nombreuses parmi les langues indo-aryennes (Martine Mazaudon, séminaire « Langues tibéto-birmanes »). A la conférence « Interfaces Prosodiques » qui s'est tenue en 2003, Carol Genetti a présenté une description du dialecte newar(i) de Dolakha (Genetti 2003, « Syntax, Prosody and Typology: Evidence from Prosodic Embedding in Dolakha Newar ») ; spécialiste de morphosyntaxe, l'auteur décrivait l'absence d'accent lexical dans la langue, et la façon dont F_0 et durée contribuent au marquage des frontières, et s'interrogeait en conclusion sur l'éventuelle existence de systèmes similaires dans d'autres langues. La réponse à la question (apportée par Philippe Martin) n'a pas tardé, signalant qu'il existait au moins une autre langue pareillement exotique, sans tons ni accentuation lexicale : le français.

3.1.3. Systèmes de tons ponctuels

La présentation des systèmes de tons ponctuels partira d'une présentation des systèmes prosodiques africains à tons ponctuels (section 3.1.3.1), pour étendre la réflexion aux systèmes asiatiques à tons ponctuels (section 3.1.3.2). Le choix d'adopter une division géographique dans la présentation est assurément contestable ; en réalité, le propos n'est pas d'opposer un « type africain » à un « type asiatique », la séparation entre ces deux ensembles ayant au contraire pour objectif de faciliter leur mise en regard.

3.1.3.1. Systèmes prosodiques africains à tons ponctuels[1]

Les systèmes prosodiques des langues à tons ponctuels d'Afrique tiennent une place de premier plan dans les perspectives typologiques esquissées plus bas (section 3.2), où sont opposés deux pôles typologiques : les *langues à intonation calculée*, et les *langues à intonation contrôlée par l'énonciateur*. Le premier de ces groupes correspond essentiellement aux langues subsahariennes.

Une brève présentation des systèmes de tons des langues africaines est proposée dans l'introduction générale à la phonologie des langues africaines par Clements 2000. En mende, exemple désormais classique de langue à tons ponctuels (Leben 1973, 1978 ; Rialland et Sangaré 1989), l'association des tons aux mots est lexicale dans le sens où chaque mot possède un schème tonal lexical ; chaque syllabe ne possède pas un ton lexical propre : l'association entre tons et syllabes est d'abord biunivoque et se fait à raison d'un ton par syllabe du début à la fin du mot. Ensuite, s'il y a des tons excédentaires, ceux-ci sont réunis sur la dernière syllabe, d'où la formation de modulations ; s'il y a des syllabes en excédent, celles-ci sont associées au dernier ton du schème tonal. Cette analyse du mende a été contestée (notamment par Dwyer 1978 et Conteh, Cowper *et al.* 1983) ; en revanche, elle s'applique de façon convaincante au kukuya (langue décrite par Paulian 1975).

Une telle association tonale au niveau du mot, qui a joué un rôle important dans le développement des premiers modèles autosegmentaux, est en fait l'exception plutôt que la règle parmi les langues africaines, qui présentent une grande variété de situations. Les caractéristiques proposés par Clements 1983 (« The hierarchical representation of tone features ») pour une typologie des langues à tons constituent autant de caractéristiques qui varient d'une langue à tons africaine à l'autre : le nombre de tons ; la présence ou l'absence de déclinaison (*declination,* parfois désignée comme *downdrift* ou *linguistically controlled downtrend*, dénominations non retenues ici) ; la présence ou l'absence de divers phénomènes de changement de plage de F_0 (tels que le *downstep* [voir définition dans le Glossaire] et inversement l'*upstep*, phénomène non décrit ici pour ne pas alourdir la discussion), et leur caractère partiel ou total. **Au-delà de sa variabilité interne, cet ensemble de langues présente plusieurs caractéristiques communes.** Au plan diachronique, les tons ne peuvent être rapportés à d'anciennes oppositions consonantiques, comme c'est le cas dans le domaine asiatique. Au stade du proto-bantu, deux tons sont reconstruits. Au plan synchronique, une caractéristique remarquable de beaucoup de langues à tons d'Afrique, celle qui a inspiré les modèles autosegmentaux-métriques de l'intonation, est que **l'action des facteurs prosodiques se traduit essentiellement par des modifications de la chaîne tonale**. Au terme de l'application de règles complexes (notamment de marquage des frontières, et d'ajustements

[1] Ce qui est exposé ici des langues d'Afrique nous est connu uniquement de seconde main, d'où un risque accru d'erreurs de fait et d'interprétation.

entre tons successifs), qui varient selon les langues, la réalisation phonétique d'un ton (par exemple, un ton L) serait la même qu'il provienne du schème tonal lexical du mot ou ait été ajouté ultérieurement au cours de la *dérivation* de l'énoncé (ou de sa *sélection*, pour employer le vocabulaire le plus récent, qui ne va pas sans soulever des difficultés ; voir Clements 2001, « Representational economy in constraint-based phonology »). **Un calcul complexe sur les séquences tonales a lieu**, dans lequel des considérations métriques entrent en jeu, cela sans doute de longue date : John Goldsmith (Goldsmith 1988, « Prosodic trends in Bantu », cité par Rialland 1995) juge que dès l'étape du proto-bantu auraient existé des propriétés de type métrique telles que l'abaissement d'une racine au ton haut après un infixe objet (infixe pronominal d'objet) au ton haut. Ces propriétés se sont développées par la suite, et diversifiées d'une langue à l'autre. Dans les langues bantu, le découpage de la phrase en groupements prosodiques est directement lié à la syntaxe : l'ordre est sujet-objet-verbe ; le sujet forme un groupe prosodique, la séquence objet-verbe en forme un second (voir par exemple Williamson 1988, « Tone and accent in Ijo » ; il semblerait que seuls le sukuma et le kikuyu fassent exception à ce constat [Cédric Patin, communication personnelle]).

3.1.3.2. Systèmes prosodiques asiatiques à tons ponctuels : implications typologiques de la description des langues à tons ponctuels d'Asie

L'étude de la place des langues à tons ponctuels d'Asie dans la typologie prosodique fournit l'occasion de mesurer la diversité des systèmes prosodiques à tons ponctuels.

Des systèmes à deux tons (H et L) sont signalés dans les langues tibéto-birmanes par Weidert 1987, et par Michailovsky 1988b et références citées (sur le cas du khaling, qui possède un ton par mot phonologique : Michailovsky 1975:206, 216). Les travaux de R.K. Sprigg, dès les années 1950, soulignent l'utilité de la notion de « registre tonal » Haut ou Bas dans la description du système prosodique du tibétain (voir Sprigg 1990). Jackson Sun propose une analyse des tons des dialectes tibétains en tons ponctuels H et L (Sun 1997). René Gsell propose de mener une analyse en tons ponctuels des contours observés dans une langue kuki-chin, le lushai (Gsell 1979a:8). Hyman et VanBik 2002b, 2002a rejoignent (indépendamment) ce point de vue : ils retrouvent dans la prosodie de plusieurs langues kuki-chin (falam, ou lai zo ; hakha ; thlatlang ; kuki-thaadow) des traits qui rappellent les systèmes prosodiques subsahariens, tels qu'ils sont décrits, en particulier, par Stewart 1962, 1993, Clements 1984 et les autres articles de l'ouvrage de Clements et Goldsmith 1984 (*Autosegmental Studies in Bantu Tone*), Hyman 1993 (« Register tones and tonal geometry »). **Il paraît utile de poursuivre la comparaison entre langues d'Asie et d'Afrique**, sans s'arrêter à la conclusion volontairement provocatrice de Larry Hyman (« les systèmes tonals kuki-chin sont 'africains' ! », conclusion d'un exposé au séminaire Langues tibéto-birmanes), car **des différences importantes apparaissent**. Cette comparaison n'en est qu'à ses débuts, notamment du fait qu'un même chercheur travaille rarement de première main sur des langues des deux continents ; elle paraît prometteuse pour une typologie affinée.

La comparaison fait ressortir que **les propriétés typiques des systèmes tonals africains** dont Clements 2000 dresse la liste **ne sont pas solidaires les unes des autres**. Cette liste comprend les phénomènes de tons *flottants* (qui causent un relèvement du ton précédent ou un abaissement des tons qui suivent) ; d'*extension*, ou *étalement*, d'un ton *(tone spreading)* sur plusieurs syllabes ; de *déplacement* de tons, qui peuvent parfois être réalisés sur des morphèmes éloignés de celui qui en est lexicalement porteur. Le naxi présente des phénomènes de tons flottants (Michaud 2003b), tandis qu'étalement tonal *(tone spreading)* et déplacement tonal y sont inconnus[1]. Une langue kuki-chin étudiée par Larry Hyman, le kuki-thaadow, connaît un phénomène de *downstep* (voir l'entrée correspondante du Glossaire) régissant la hauteur phonétique respective des tons H qui se succèdent dans l'énoncé ; ce phénomène est absent d'autres langues kuki-chin : falam (lai zo), hakha, thlanglang. (Des situations semblables s'observent dans le domaine bantu : de deux langues proches parentes, l'une peut présenter un *downstep* et l'autre non.)

Inversement, certains traits peuvent être partagés par des langues dont les tons lexicaux de base sont profondément différents. **Les phénomènes de réassociation tonale, tels qu'il en existe en naxi, ne sont pas particuliers aux langues à tons ponctuels.** Des phénomènes de sandhi et de modification tonale catégorielle sont présents dans certaines langues de la famille hmông-miên, dont les tons modulés sont par ailleurs proches de ceux de langues comme le vietnamien, au plan synchronique comme au plan diachronique (Niederer 1998, et recherches en cours de D. Mortensen ; au sujet des tons modulés, voir ci-dessous, section 3.1.5). Certains dialectes chinois présentent des tons flottants : en chinois cantonais, certains tons au contour moyen-montant proviennent de syllabes (sémantiquement parentes) possédant un ton plain (Kam Tak Him 1977) ; le cantonais possède également un processus productif de réassociation d'un ton haut d'un mot grammatical (dont tous les segments disparaissent) vers la syllabe qui le précède, créant pareillement un contour montant (Hashimoto-Yue 1987, « Tone sandhi across Chinese dialects » ; Chen 2000:60 ; Yu 2003, « Some methodological issues in phonetic typology research: Cantonese contour tone revisited »).

Les changements tonals naxi d'origine morphosyntaxique (par réduction syllabique) et intonative (par un processus catégoriel d'intensification intonative), évoqués au chapitre II, section 2.4, ont de proches équivalents dans d'autres langues d'Asie comme le bai. La langue bai, voisine de la langue naxi, a un système tonal profondément différent, certains tons étant caractérisés par une phonation *tendue* (Xu Lin et Zhao Yansun 1984 ; Edmondson, Esling *et al.* 2001 ; Esling et Edmondson 2002) ; cette langue n'en possède pas moins deux phénomènes catégoriels de changement tonal qui rappellent le naxi : création

[1] Dans une autre langue tibéto-birmane, le maru, il existe trois tons lexicaux de base H, M, L, comme en naxi, et un ton H flottant morphologique (Okell 1988) ; à la différence du naxi, ce ton flottant n'a pas pour effet la création de contours, mais le rehaussement de L en M et de M en H.

d'un ton modulé par un processus catégoriel d'intensification, qui n'affecte que certains tons ; et création d'un ton modulé par réduction syllabique[1]. Dans le dialecte de bi[4]jiang[1] 碧江, un ton descendant peut devenir montant sous l'effet d'un processus d'intensification[2]. Le processus de réduction syllabique s'observe dans le dialecte de jian[4]chuan[1] 剑川. Certains mots polysyllabiques portant un ton /[21]/, /[33]/ ou /[44]/ sur leur première syllabe et un ton haut (/[55]/) sur leur deuxième syllabe peuvent perdre cette deuxième syllabe ; son ton haut s'associe alors à la première syllabe, ce qui aboutit à un contour montant /[35]/, quel que soit le ton lexical porté à l'origine par la syllabe initiale de mot. Par exemple, /tɯ[21] kɯ[55] tsi[55]/ (« autrefois, par le passé ») devient /tɯ[35] tsi[55]/ (Xu Lin et Zhao Yansun 1984:8).

Il n'est pas exclu que ces phénomènes soient passés d'une langue vers l'autre, dans des conditions de bilinguisme relativement étendu. (L'étude des contacts passés entre le naxi et les langues voisines n'en est qu'à ses débuts.) **L'importance de la convergence aréale ne paraît pas moindre au plan de l'accentuation lexicale (accent, accents mélodiques ou tons) qu'au plan phonémique.** Svantesson 2001 (« Tonogenesis in Southeast Asia: Mon-Khmer and beyond ») attribue à l'influence de langues tonales voisines le développement de tons en hu (langue angkuique parlée dans le Yunnan) ; cette langue ne possédait pas les oppositions consonantiques dont la transphonologisation donne naissance à des oppositions de qualité de voix puis aux tons dans le schéma classique du développement tonal (décrit ci-dessous, sections 3.1.5.2-3) ; sous la pression aréale vers le développement de tons, elle a suivi un chemin original : la transphonologisation d'une différence de longueur vocalique en tons (voyelle longue donnant ton bas, voyelle brève ton haut). La langue u, dans une situation similaire, aurait partiellement transphonologisé l'opposition de hauteur vocalique entre i ~ e et o ~ u en opposition tonale (Svantesson 2001)[3].

3.1.4. Registres de qualité de voix

Le terme de « registre » est ambigu, étant parfois employé comme synonyme de « niveau de hauteur mélodique », soit de façon informelle (remarquant qu'une syllabe est réalisée *dans un registre élevé*, ou *dans un registre grave*, sans que cela implique une division

[1] Les faits exposés ici proviennent de l'ouvrage de Xu Lin et Zhao Yansun 1984:7 ; ils ont été confirmés par le Pr. Xu Lin (communication personnelle).

[2] Le ton de départ est soit le ton /[42]/, soit le ton /[21]/ ; dans cette notation à cinq degrés, créée par Chao Yuen-ren, 5 désigne le haut de la plage de fréquence fondamentale du locuteur, 1 le bas ; les successions de chiffres indiquent un mouvement mélodique d'un niveau de hauteur vers un autre. L'aboutissement du processus est un ton /[13]/.

[3] De même, dans le domaine des langues à accent, l'influence des langues germaniques sur le français aurait eu des conséquences étendues sur la structure syllabique de la langue et son évolution segmentale : Vaissière 2001, « Changements de sons et changements prosodiques: du latin au français »). Dans les dialectes basques, le continuum entre accent mélodique et accent lexical proche de celui de l'espagnol a une distribution géographique qui suggère une influence de l'espagnol (Hualde, Elordieta *et al.* 2002, « From pitch accent to stress accent in Basque »).

catégorielle), soit pour désigner des niveaux discrets, dans la description des langues à tons ponctuels (ainsi, les trois principaux tons du naxi pourraient être décrits comme : registre haut, registre moyen, registre bas). ***Langue à registres*** **est employé ici dans le sens de *langue à registres de qualité de voix* ; les langues qui opposent deux ou plusieurs niveaux de hauteur tonale (« langues à registres tonals ») sont appelées *langues à tons ponctuels*** (pour plus de précisions, voir le Glossaire).

Les langues à registres de qualité de voix font un usage lexical distinctif de la qualité de voix : selon les cas, voix soufflée *(breathy voice)*, voix tendue *(tense voice)*, et/ou voix laryngalisée *(creaky voice)* s'opposent à une qualité de voix non marquée, communément désignée comme *voix modale*. Un panorama des emplois linguistiques de la qualité de voix est proposé par Gordon et Ladefoged 2001, qui dressent la liste de 45 langues employant des oppositions de qualité de voix.

Il est désormais bien établi que les registres de qualité de voix sont un stade intermédiaire dans la transphonologisation des oppositions consonantiques en oppositions tonales ou en oppositions vocaliques[1]. Il paraît donc plausible d'affirmer que la plupart des langues à tons d'Asie ont connu, à un moment ou un autre de leur histoire, un stade auquel elles opposaient des registres de voix. Un travail fondateur dans ce domaine est l'article d'Eugénie Henderson sur le cambodgien (Henderson 1952, « The main features of Cambodian pronunciation » ; voir également Ferlus 1992) : l'orthographe de cette langue témoigne d'une ancienne distinction entre initiales voisées et non voisées ; en cambodgien moderne, à partir du XVI^{e} siècle, les initiales voisées et non voisées se sont confondues en non voisées, cependant que l'opposition entre initiales se transphonologisait en une opposition entre phonation soufflée *(breathy voice)* sur les syllabes à anciennes initiales voisées, et voix modale sur les autres. Ce trait n'est plus observable que dans certains dialectes conservateurs, l'opposition ayant évolué vers une opposition vocalique (bipartition du système vocalique) dans les principaux dialectes ; Eugenie Henderson a eu la bonne fortune de travailler avec deux consultants dont l'un conservait une distinction entre registres de qualité de voix tandis que l'autre ne faisait guère plus la distinction que par les voyelles :

> Mr. Keng, as a philosophy student with literary and dramatic leanings, was aware of and interested in language from both the philosophic and æsthetic standpoints. His style of utterance was in general more deliberate and controlled than that of Mr. Mongkry, who as a student of economics was less concerned with language for its own sake. The two styles complemented each other well. Mr. Keng's style was

[1] Ainsi, l'étude récente d'Abramson, Thongkum *et al.* 2004, « Voice register in Suai (Kuai): An analysis of perceptual and acoustic data », qui visait à décrire au plan phonétique la réalisation du registre soufflé *(breathy)* de ce parler, aboutit au constat que le développement de tons distinctifs est déjà très avancé dans le dialecte considéré, remplaçant l'opposition de registres de qualité de voix. La raison pour laquelle la transphonologisation se fait en direction de tons plutôt que de contrastes vocaliques tient très certainement à l'influence de la langue nationale, le thai.

> helpful in that the different voice quality and manner of utterance of the two registers were clearly, sometimes startingly, recognizable, even in fairly rapid speech, whereas Mr. Mongkry appeared often to make no distinction other than that of vowel quality. On the other hand, Mr. Mongkry's style of utterance was valuable for the ease and naturalness with which the alternative pronunciations proper to rapid speech were forthcoming. (p. 149)

Eugenie Henderson a ainsi pu saisir dans sa dynamique la transphonologisation des registres en oppositions vocaliques. Sa description synchronique des registres est la suivante :

> The characteristics of the first register are a 'normal' or 'head' voice quality, usually accompanied by relatively high pitch.
> The characteristics of the second register are a deep rather breathy or 'sepulchral' voice, pronounced with lowering of the larynx, and frequently accompanied by a certain dilation of the nostrils. Pitch is usually lower than that of the first register in similar contexts.

La langue khmu (kammu) présente une variété de dialectes exemplaire du phénomène : ces dialectes par ailleurs très proches sur de nombreux plans (inventaire phonémique, morphologie, morphosyntaxe, lexique) varient par leur accentuation lexicale[1] : certains conservent des oppositions de voisement à l'initiale, tandis que d'autres les ont perdues et possèdent, selon les cas, des oppositions tonales (ton bas après les initiales anciennement voisées, ton haut ou descendant après les initiales anciennement non voisées) ou de qualité de voix[2].

Une procédure expérimentale pour décider si une langue doit être classée dans le groupe des langues à registres ou dans le groupe des langues à tons est proposée par Gandour, Gårding *et al.* 1978 et Svantesson et House 1996 : ces auteurs jugent que si la manipulation de F_0 sur une même syllabe permet de synthétiser des stimuli qui couvrent toutes les catégories que distingue la langue, et soient catégorisés comme tels par les auditeurs, le système est tonal. (Voir également Suwilai 2003, et les commentaires de Svantesson : Svantesson 2003.) Ce critère conduirait par exemple à conclure qu'en chinois mandarin, la laryngalisation relativement fréquente du troisième ton est un trait secondaire, non définitoire du ton, si elle n'est pas indispensable à la reconnaissance de ce ton. L'application du test soulève des questions techniques : la variation simulée de F_0 doit correspondre à ce qu'un larynx humain peut produire en voix modale ; il importe de vérifier que la synthèse ne produit pas des *monstres :* des stimuli qui n'auraient pu être produits par un larynx humain, et donnent une impression auditive dont l'équivalent le plus proche dans la parole naturelle est produit par

---

[1] Rappelons que le terme d'« accentuation », dans le vocabulaire proposé par J. Vaissière, recouvre accent distinctif, tons, accents mélodiques, et également registres de qualité de voix.

[2] Svantesson 2003 juge que certains dialectes khmu seraient passés d'une opposition de voisement à une opposition tonale haut-bas sans connaître le passage par l'étape intermédiaire des registres de voix ; si elle est confirmée, cette observation pourrait apporter un éclairage nouveau sur la question.

un changement de qualité de voix. Pour prendre un exemple caricatural, l'abaissement soudain de fréquence fondamentale à la fin du ton B2 vietnamien est la conséquence d'une constriction glottale ; calquée en synthèse de la parole, sa courbe de F_0, sans équivalent ailleurs dans le système tonal vietnamien, pourrait sans doute donner aux auditeurs l'impression du ton B2 ; doit-on en tirer argument pour décrire le ton B2 comme un ton mélodique, dans lequel la qualité de voix joue un rôle secondaire ? Le critère employé par Gandour, Gårding *et al.* 1978 et Svantesson et House 1996, décisif pour comparer le rôle de F_0 dans divers dialectes d'une même langue (en employant les mêmes stimuli), ne permet sans doute pas pour autant de tracer de ligne de démarcation claire entre langues à registres et langues à tons ; le passage du premier type au second paraît être un processus scalaire. **Les types mixtes sont nombreux** (particulièrement dans les langues austro-asiatiques) : **langues conservant un certain usage de la qualité de voix à un stade où les oppositions sont déjà réalisées et perçues à l'aide de la mélodie**. Matisoff 1999:15 propose la formule de « phonation-prominent tone systems », prenant pour exemple le birman de Rangoon ; ces cas intermédiaires, notamment celui du tamang (Mazaudon 1973), seront évoqués ci-dessous, section 3.1.5.

Enfin, il paraît intéressant de noter que **les registres ne sont pas exclusifs d'un emploi intonatif de la qualité de voix**. Eugenie Henderson rapporte l'observation suivante :

> In sentences the word registers are modified according to intonation and by emotional factors. Register may be used, as in many other languages, to express emotion, and when this happens the emotional register may overlie the lexical register, much as in many tone-languages intonation may overlie lexical tone. (Henderson 1952)

Cette formulation s'applique aussi bien aux faits vietnamiens exposés au chapitre III : l'emploi lexical de la qualité de voix limite sa variation intonative, sans la supprimer ; **les traits lexicaux fournissent le cadre dans lequel l'écart intonatif prend son sens**.

3.1.5. Tons modulés

3.1.5.1. Le débat sur l'existence de tons modulés. Les propositions de réanalyse des tons modulés en tons ponctuels

La distinction entre langues à tons ponctuels et langues à tons modulés (« gliding-pitch contour systems ») remonte à Kenneth Pike. La distinction entre les deux ensembles est la suivante : « the glides of a contour system must be treated as unitary tonemes and cannot be broken down into end points which constitute lexically significant contrastive pitches » (Pike 1948:10). Pike souligne la présence de types intermédiaires, tout en maintenant la nécessité de distinguer les deux catégories.

La distinction correspond en partie à un découpage géographique : les langues à tons africaines et amérindiennes sont majoritairement des langues à tons ponctuels ; les langues à tons modulés les plus étudiées appartiennent à la famille austroasiatique et à la famille sino-tibétaine. Pour autant, **les langues d'Asie comptent aussi des langues à tons ponctuels, dont le naxi et les langues kuki-chin** (voir ci-dessus, section 3.1.3.2). René Gsell remarque :

> Des travaux récents de tonologie, en partie d'inspiration générativiste, ont montré la nécessité d'une analyse en traits distinctifs des tonèmes d'une langue à tons. (Gsell 1979b:29)
>
> En Asie, il ne semble pas que l'interprétation en mores successives de tons modulés pose des problèmes insolubles (...). Nous avons longuement discuté ces problèmes dans nos travaux sur le thai et le vietnamien et conclu qu'en thai, il y avait deux niveaux pertinents (haut et bas), plus un niveau non marqué qui forme le ton moyen, les tons modulés sont des successions de mores de niveaux distinctifs différents. (Gsell 1979a:7)

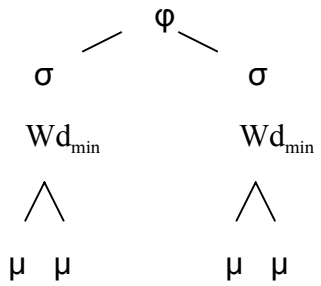
La formulation de René Gsell revient à contester l'existence de tons modulés qui ne soient pas décomposables en tons ponctuels : il serait nécessaire d'étendre à l'ensemble des langues à tons une analyse en tons ponctuels et en *mores tonales*[1]. Une telle extension serait satisfaisante au plan de l'économie descriptive, car elle permettrait d'opérer avec les unités proposées au statut d'universaux du langage par le courant de la phonologie générative (de la plus petite unité à la plus grande : more ; syllabe ; pied ; mot prosodique ; groupe phonologique ; groupe intonatif ; énoncé), tandis que la reconnaissance de tons modulés non analysables suscite la crainte d'une prolifération des catégories linguistiques[2]. Il existe

---

[1] Il est même imaginable qu'un découpage en mores rende des services dans la description de la qualité de voix, prenant exemple sur l'analyse du *stød* danois (qui se traduit au plan phonétique par une glottalisation) par Basbøll 1999, Grønnum et Basbøll 2001 (« Consonant Length, Stød and Morae in Standard Danish ») : ces auteurs proposent une analyse dans laquelle le *stød* signale la deuxième more d'une syllabe, fonctionnant en cela comme compteur de mores (Grønnum et Basbøll 2001:234).

[2] La préparation d'un mémoire de maîtrise portant sur les théories de la référence au début du vingtième siècle (*A Comparative Study of Theories of Reference: Meinong, Husserl, Frege*) nous a permis de mesurer l'attachement de la grande majorité des auteurs anglo-saxons (Quine, Russell, Leonard Linsky, P.F. Strawson, plus récemment David Bell, Michael Dummett) aux préceptes radicaux un moment défendus par Wittgenstein concernant l'*économie dans les concepts employés*, qui se reflète dans un attachement à des oppositions binaires telles que : une proposition est vraie ou non ; un objet existe, ou pas. Ces auteurs adoptent la formule de Guillaum d'Ockham (XIIIe siècle), « Non sunt multiplicanda entia praeter necessitatem » (que l'on peut traduire par « il ne faut pas multiplier les êtres sans nécessité »), qu'ils entendent dans un sens strictement nominaliste, au risque d'oublier le *praeter necessitatem* et d'ériger la *politique d'austérité* en qualité philosophique. Cette orientation les conduit à rejeter sans appel les penseurs dits « réalistes » dont Meinong (*Über Annahmen*, 1910, Verlag von Johann Ambrosius Barth, Leipzig, réimpr. 1968-78 : *Alexius Meinong Gesamtausgabe*, Akademische Druck- und Verlagsanstalt, Graz, édition dirigée par R. Haller, R. Kindinger et R.M. Chisholm), qui par souci de vérité psychologique reconnaissent le statut d'*objet de la pensée* à égalité aux objets réels et aux objets de fiction, licornes et chimères. La *théorie des objets* de Meinong distingue nettement

diverses analyses en tons ponctuels des tons des dialectes chinois et d'autres langues de la région, notamment celles de Moira Yip (Yip 1980, 1992, 1993, 2002) et Duanmu San (Duanmu San 1990, 2000). « The prosodic hierarchy of Chinese includes the familiar entities mora (μ), syllable (σ), foot (φ). In many languages the notion minimal word (Wd_{min}) also plays a role, and typically can be equated with the foot. In Chinese languages it is more usually to be equated with the syllable. » (Yip 1992)



L'idée selon laquelle les tons de toute langue sont analysables en combinaisons de tons H et L est si répandue que l'idée contraire suscite aujourd'hui la surprise chez des collègues phonéticiens-phonologues non familiers des langues à tons d'Asie. Pourtant, **les analyses en mores tonales qui ont été proposées dans le domaine chinois n'emportent pas nécessairement l'adhésion des spécialistes**. Ainsi, Larry Hyman (communication personnelle) note que, bien qu'il soit théoriquement possible d'analyser les tons chinois en termes de séquences de tons ponctuels, cette réanalyse n'est pas éclairante en pratique : contrairement à ce qui serait attendu s'il s'agissait bien, à un niveau sous-jacent, de séquences de tons ponctuels, les phénomènes de sandhi observés n'ont pas pour effet de simplifier ces « séquences ». (Voir également les critiques de Dell 2004 à l'endroit des travaux de Duanmu San.) **L'économie conceptuelle que représente la réduction des contours à des séquences de tons ponctuels a pour prix la multiplication d'entités d'un autre ordre** : *mores tonales* et *séquences de tons associées à une même syllabe*, entités qui ont fait la preuve de leur utilité dans la description de certaines langues (voir notamment Poser 1990, Nagano-Madsen 1992), mais dont l'utilité ne s'impose pas dans la description (diachronique et synchronique) de beaucoup de langues à tons d'Asie. Dans la discipline largement empirique qu'est la linguistique, la valeur des propositions se mesure par leur pouvoir d'explication et de prédiction ; le découpage de la syllabe en mores ne paraît pas avoir fait les preuves de son utilité dans l'analyse du chinois mandarin (non plus que du vietnamien et d'autres langues du domaine austroasiatique).

Le paragraphe 3.5.1.2 évoque certains des travaux qui portent sur l'apparition et l'évolution des tons modulés d'Asie ; le paragraphe 3.5.1.3 présente en détail l'exemple du vietnamien.

---

les objets du monde intérieur, corrélats d'opérations mentales, et les objets du monde extérieur ; pour cela, il divise la notion d'*existence* en plusieurs sous-catégories. De tels anathèmes paraissent regrettables ; de même, en linguistique, la tension permanente vers l'économie des concepts, dans le projet d'une grammaire universelle, porte pareillement le risque d'une exclusion d'idées en fait utiles, et d'un traitement réducteur de la réalité qu'il s'agit de comprendre.

3.1.5.2. Apparition et évolution des tons modulés

Des travaux abondants décrivent la dynamique diachronique des tons modulés d'Asie, typiquement issus de la perte d'oppositions consonantiques, *via* un stade où la distinction est portée par la qualité de voix (voir en particulier Haudricourt 1954, 1961, 1965 ; Henderson 1952, 1967 ; Ferlus 1979, 1998b, 2001 ; Matisoff 1970, 1973b, 1999 ; Mazaudon 1977 ; Sagart 1988 ; Niederer 1998). Les travaux d'orientation diachronique sont complétés par des travaux expérimentaux (depuis Liu Fu 1925, *Etude expérimentale sur les tons du chinois*), en particulier Huffman 1987 ; Svantesson et House 1996 ; Svantesson 2001, 2003 ; Abramson, Thongkum *et al.* 2004 ; Edmondson, Esling *et al.* 2001 ; Nguyên Van Loi et Edmondson 1998 ; Wayland et Jongman 2003. Certaines hypothèses peuvent être avancées au sujet de la réalisation phonétique des tons lors de leur apparition en chinois et en vietnamien : lors de la bipartition, la série voisée donne des tons *de série basse*, la série non voisée des tons *de série haute* (dans les rares cas où la bipartition est conditionnée par l'aspiration, les non aspirées forment la série haute, les aspirées la série basse ; voir par exemple Sagart 1993:111). Par la suite, en revanche, **les tons connaissent une évolution qui peut les conduire très loin de leur valeur d'origine**. Laurent Sagart résume ainsi le modèle de Hirayama 1984, que ses propres observations confirment :

> « ... les contours des tons des dialectes chinois évoluent selon un schéma circulaire relativement fixe : haut égal → tombant → bas égal ou convexe → montant → haut égal. Des propositions similaires ont été faites indépendamment par d'autres auteurs pour expliquer les variations de contours de l'accent mélodique dans les dialectes suédois (Bruce et Gårding 1978) et dans les langues baltes (Martinet 1986:75). » (Sagart 1993:116)

A notre connaissance, cette dynamique n'a pas d'équivalent dans les langues subsahariennes.

3.1.5.3. Un exemple développé de langue à tons modulés et qualité de voix : le vietnamien

Le vietnamien apparaît comme un exemple révélateur, à la fois par la complexité de son système tonal, et par le rôle qu'a eu cette langue dans la découverte des mécanismes d'apparition des tons. Il peut être utile de récapituler brièvement les étapes de l'apparition des tons vietnamiens.

Historiquement, le proto-viet-muong était une langue non tonale qui possédait des syllabes ouvertes, des syllabes qui se terminaient par une constriction glottale / ˀ / (le symbole / ˀ / ne figure pas dans l'alphabet phonétique international ; il remonte à Smalley 1963:389ff), des syllabes qui se terminaient par un /h/, et enfin des syllabes fermées par une consonne occlusive orale ; voir le tableau 4.2a (adapté de Ferlus 2001:2). (Laurent Sagart suggère que le /h/ était en réalité un /s/ : Sagart 1988.) L'opposition entre les trois premiers types de syllabes s'est transphonologisée, *via* un stade où elle était portée par la qualité de voix

(Sagart 1988, Abramson 2004), en opposition entre trois tons lexicaux (perte des oppositions segmentales et développement concomitant de tons, préservant les distinctions lexicales).

> ... les laryngales sont les seules [consonnes] dont l'articulation ne fait intervenir que la glotte, plus exactement les muscles et les parois de la glotte. Or les manifestations de variation mélodique, intonation ou ton, elles aussi ne font intervenir que ces mêmes organes. Il n'est donc pas surprenant qu'une modification des laryngales finales, amüissement suivi de chute, entraîne l'apparition de traits mélodiques. (Ferlus 1998b:21)

Les trois tons résultants sont communément désignés comme A, B et C (suivant l'emploi de Gedney 1989), notation abstraite qui ne dit rien de leur réalisation phonétique. A ce stade (tableau 4.2b), il existait donc trois tons distinctifs sur les syllabes à sonante finale, et les syllabes à occlusives finales formaient un ensemble distinct, à l'intérieur duquel n'existait pas de différenciation tonale. Plus tard, l'opposition entre initiales voisées et non voisées s'est également transphonologisée en opposition tonale (avec un stade intermédiaire où l'opposition était portée par la qualité de voix soufflée, *breathy*, de la voyelle : Ferlus 1998b:6), occasionnant une bipartition du système tonal (Haudricourt 1961) qui aboutit au paradigme actuel de six tons en syllabes à finale sonante (tons A1 à C2) et deux tons pour les syllabes à occlusives finales (tons D1 et D2), comme le montre le tableau 4.2c.

ta	taˀ	tah
da	daˀ	dah

tap, tat, tak
dap, dat, dak

Tableau 4.2a : types de rimes en proto-viet-muong tardif (Late Proto-Viet-Muong) : *syllabes ouvertes non glottalisées, syllabes finissant sur une constriction glottique, syllabes à finale /h/, et syllabes à finale /p/ /t/ /k/.*

ta	ton A	ta	ton B	ta	ton C
da		da		da	

tap, tat, tak	catégorie D
dap, dat, dak	(pas de ton)

Tableau 4.2b : première étape de la tonogenèse en vietnamien : trois tons ; pas de ton distinctif sur les syllabes à obstruantes finales.

ta	ton A1 (ngang)	ta	ton B1 (sắc)	ta	ton C1 (hỏi)
ta	ton A2 (huyền)	ta	ton B2 (nặng)	ta	ton C2 (ngã)

tap, tat, tak	ton D1 (sắc)
tap, tat, tak	ton D2 (nặng)

Tableau 4.2c : le système tonal du vietnamien contemporain (dialecte de Hanoi) : numéro des tons et nom du ton dans l'orthographe actuellement en usage. Les tons A1 à C2 apparaissent seulement sur les syllabes finissant par une sonante (voyelle, ou consonne nasale /m/, /n/, /ŋ/) ; les tons D1 et D2 apparaissent seulement sur les syllabes à obstruente finale.

Tableau 4.2. Les tons vietnamiens en perspective diachronique, d'après Ferlus 2001:2.

Une représentation schématique des réalisations phonétiques des huit tons du vietnamien d'après des données du locuteur M1 est fournie par la figure 4.1. **Les tons A1, A2 et B1 sont produits en voix modale** : A1 est haut-plain, A2 bas-descendant, B1 haut-montant. **Les tons B2 et C2 sont glottalisés** : le ton B2 comporte une constriction glottale croissante du début de la rime jusqu'à l'interruption du voisement ; le ton C2 comporte une constriction médiane et s'achève sur une fréquence fondamentale plus élevée qu'en début de rime. **Le ton C1 est produit avec un voisement typiquement plus relâché** ; il varie (entre locuteurs, et parfois chez un même locuteur) entre une réalisation haute-descendante-montante et une réalisation descendante avec relâchement laryngal final (se traduisant par une laryngalisation, ou une forte friction, comme une illustration extrême de la voix soufflée ; voir Vu-Ngoc, d'Alessandro *et al.* 2005). Thompson 1965 fournit une description non instrumentale très précise des six principaux tons ; des données expérimentales sont apportées par Pham Thi Ngoc Yên, Castelli *et al.* 2002:26 ; Vu Ngoc, d'Alessandro *et al.* 2002 ; Pham 2003b.

En vietnamien, le ton est associé à la syllabe, non à une unité de taille supérieure ; les changements catégoriels de tons sporadiquement observés appellent une explication en termes d'évolution diachronique des syllabes concernées, non en termes de changements syntagmatiques *(sandhi tonal).* Ainsi, le nombre « dix » a deux réalisations tonales, selon qu'il apparait dans les nombres de 10 à 19 ou dans les nombres de 20 à 99 : « dix » est *mười* (notation API : /muɤj/ ; ton : A2, ton bas) ; « vingt » est *hai mươi* (mêmes phonèmes que précédemment : /muɤj/, tandis que le ton devient un ton haut : A1). D'autres paires de mots, l'un au ton A1 et l'autre au ton A2, retiennent l'attention : « beaucoup » se dit *nhiều*, syllabe au ton A2 ; « combien » se dit *bao nhiêu*, la syllabe *nhiêu* étant au ton A1. Ces cas appellent en réalité un traitement similaire à celui de : *ngồi* « s'asseoir » et *ngôi* « introniser » ; *voi* « éléphant » et *vòi* « trompe » : il ne s'agit pas de sandhi, mais de l'influence d'une autre syllabe (forme de préfixe) présente lors des dévoisements qui ont abouti à la bipartition du système tonal. Il est imaginable qu'il y ait eu une syllabe à initiale non voisée avant le nombre « dix » dans l'expression qui est devenue *hai mươi*, « vingt » : vingt aurait été « deux dizaines », trente « trois dizaines », la différence entre dix et dizaine se marquant par une syllabe additionnelle, dont il ne reste plus trace aujourd'hui que dans le ton de la syllabe *mươi*. Cette syllabe-préfixe aurait eu une initiale non voisée, commandant l'entrée en catégorie haute (ton A1) lors de la bipartition tonale. Le détail de l'analyse dans ces cas précis reste à mener ; les phénomènes de cet ordre sont en revanche bien attestés dans les langues de la région[1].

3.1.5.4. Le domaine du ton n'est pas nécessairement la syllabe

Même en ne prenant en compte que les langues à tons d'Asie, le groupe des langues à tons modulés présente une grande diversité interne. Dans certaines langues, que Matisoff appelle

[1] Merci à Michel Ferlus pour d'instructives discussions au sujet de cette question.

« omnisyllabiques », chaque syllabe porte un ton (par exemple en naxi) ; mais **le domaine du ton est parfois le *mot phonologique* dans son entier**. Ainsi, dans les langues parlées au Népal qui appartiennent à la branche *bodish* du tibéto-birman, c'est généralement le mot phonologique qui est le domaine des oppositions tonales (Mazaudon 1977:81). Le tamang, langue tibéto-birmane parlée au Népal, possède quatre tons qui se déploient sur le mot entier, et non sur une seule syllabe :

> Phonétiquement (...) les syllabes non initiales ne présentent ni une répétition du ton précédent, ni une réalisation spéciale constante, ni une variation libre, mais varient en fonction du ton du lexème et de leur propre position par rapport au début et à la fin du mot, de manière à supporter une partie de la courbe caractéristique du ton du lexème. (Mazaudon 1973)

La situation du tamang est donc nettement différente de celle de langues à schème tonal lexical comme le mende, du fait que **le déploiement du ton sur plus d'une syllabe ne peut se décrire en termes d'une répartition de plusieurs tons entre les syllabes**. Le tamang se distingue également des langues à accent mélodique, par le fait que les quatre tons s'opposent sur les monosyllabes, tandis que dans une langue comme le suédois, la distinction portée par l'accent mélodique demande au moins deux syllabes pour se réaliser (fait notamment relevé par Martinet 1960). « Le tamang constitue un cas intermédiaire entre les langues qui ont une opposition tonale par syllabe, et celles où une seule syllabe ne saurait suffire au développement d'une opposition tonale » (Mazaudon 1973:92).

Le système tonal tamang est relativement complexe au plan phonétique : même lorsque les mots sont élicités à l'intérieur d'une phrase-cadre, les courbes de F_0 présentent une forte marge de variation, et ne sont pas séparées par un « intervalle de confiance » (formule employée par René Gsell au sujet du vietnamien : Gsell 1980) ; Martine Mazaudon rapporte cette variabilité à une variation concomitante de qualité de voix, formulant l'hypothèse d'une relation de compensation entre degré de *souffle (breathiness)* et hauteur, la réalisation de la consonne participant également à la réalisation des oppositions tonales. (Une vérification expérimentale est en cours[1].)

[1] Les premiers résultats de l'étude acoustique et électroglottographique en cours ont été présentés au 11th Himalayan Languages Symposium, Université de Chulalongkorn, Bangkok (6-9 déc. 2005), sous le titre « Issues in the voice quality of Tamang tones: an electroglottographic study » (M. Mazaudon, A. Michaud).

Bilan

Dans le domaine de l'analyse des tons lexicaux comme dans celui des études intonatives, certaines barrières se dressent entre chercheurs d'orientations opposés : les tenants de l'analyse en tons ponctuels sont soupçonnés d'impérialisme (notamment du fait de la généralisation des symboles des tons ponctuels dans la version la plus récente de l'Alphabet Phonétique International, qui revient à encourager une analyse en tons ponctuels) ; pour leur part, ils tendent à juger que le fond n'a pas encore été atteint dans l'analyse des tons mélodiques dits *non décomposables*. La chance typologique de la présente étude est qu'elle repose sur l'analyse d'une langue à tons ponctuels (le naxi) et d'une langue à tons complexes (le vietnamien), ce qui, espérons-nous, lèvera le soupçon de partialité en faveur de l'un ou l'autre camp. Les observations effectuées, jointes à un aperçu typologique, vont dans le sens des observations de Pike : tons ponctuels et tons modulés paraissent représenter deux pôles non réductibles l'un à l'autre. En revanche, l'un et l'autre groupe présentent une grande variabilité interne.

3.1.6. Entre système tonal et système accentuel : la question de l'accent dans les systèmes tonals, et les systèmes à accent mélodique (pitch accent)

3.1.6.1. Problématique générale : tons et accent

André Martinet note qu'une même langue peut combiner tons et accents :

> Il est possible, pour une langue, de combiner tons et accents, mais il semble que, dans ce cas, les différences tonales ne puissent se maintenir que sous l'accent. Le grec classique combinait un accent libre (sur une des trois dernières syllabes du mot, mais avec des restrictions) avec deux tons appelés traditionnellement aigu et circonflexe. Le letton distingue trois tons qui se manifestent dans la syllabe initiale qui porte l'accent fixe. (Martinet 1956:49 ; voir aussi Trubetzkoy 1939 [1976]:226)

En angami, langue tibéto-birmane qui possède cinq tons et un accent lexical, une neutralisation partielle des contrastes tonals et vocaliques a lieu en syllabe inaccentuée : seuls deux des cinq oppositions tonales sont conservées en position inaccentuée[1]. Sur la base d'observations semblables, Larry Hyman propose les types suivants (tableau 4.3) :

	accent	**pas d'accent**
ton	fasu (námo mē ; langue de Papouasie-Nouvelle Guinée), suédois, palantla chinantec (langue parlée au Mexique)	yorùbá, igbo
pas de ton	anglais, russe	coréen, berbère, gafat (langue d'Ethiopie)

Tableau 4.3. Principaux types de systèmes prosodiques, d'après Hyman 2001c.

[1] Merci à Martine Mazaudon pour nous avoir opportunément signalé l'exemple de l'angami.

Les paragraphes qui suivent discutent le statut de l'accent et des syllabes non accentuées dans des langues à tons.

L'exemple du chinois mandarin standard : les syllabes « inaccentuées » sont beaucoup moins fréquentes que les syllabes « accentuées »

En mandarin standard, l'extrême majorité des syllabes porte l'un des quatre tons lexicaux distinctifs (traditionnellement illustrés par la syllabe /ma/ avec les tons 1 à 4 : 妈，麻，马，骂, qui signifient respectivement « maman », « chanvre », « cheval », « insulter ») ; certaines syllabes portent un « ton léger » (qing^{1}sheng^{1} 轻声), aussi appelé « ton neutre » (Chao Yuen-ren 1968) ou « ton zéro ». Ces dernières apparaissent sur la deuxième syllabe de certains disyllabes. Le phénomène est lexical, au sens où il est possible de proposer des paires minimales telles que : wen^{2}zi^{0} 蚊子 « moustique », wen^{2}zi^{4} 文字 « caractère, signe écrit ». La syllabe au ton léger est souvent une forme de suffixe, typiquement zi^{3} 子 « enfant » (devenant zi^{0}) ou tou^{2} 头 « tête » (devenant tou^{0}) (des observations similaires sont rapportées dans divers dialectes chinois ; au sujet des dialectes gan, voir Sagart 1993:38). Le phénomène est relativement fluctuant : les dictionnaires ne s'accordent pas toujours sur le statut du ton des syllabes concernées, ton « plein » ou ton « léger » (Hoa 1980:258). Il s'agit d'un phénomène relativement marginal, qui rappelle l'affaiblissement de certains mots grammaticaux (的，了，地，么), dont la voyelle est réduite et qui ne possèdent plus de ton propre. « Weak stress » et ton léger nous paraissent interchangeables dans l'emploi qu'en fait Chao Yuen-ren 1948. Autrement dit, toutes les syllabes sont lexicalement « accentuées », à l'exception des syllabes au ton léger, qu'il est possible de déclarer « désaccentuées ». (Une conclusion similaire ressort de la discussion par Chan 1993:346 du livre de Shen 1990.) De la sorte, le mandarin standard combine tons et accent (comme le remarquent, entre autres auteurs, Martinet 1956:49 et Jun 2005:431), mais à la différence du système accentuel de langues sans tons (comme le russe), la catégorie *marquée* est la catégorie des syllabes non accentuées.

Les dialectes wu du chinois présentent des phénomènes de type accentuel plus développés que le mandarin standard ; l'accent y est solidaire d'une neutralisation tonale sur les syllabes non accentuées (voir notamment Chen 2000). Le troisième grand ensemble dialectal, le cantonais, a des tons « omnisyllabiques », selon la formule de James Matisoff : chaque syllabe possède un ton propre ; il n'existe pas de « ton léger ».

Réflexion générale sur les syllabes « inaccentuées » (ou syllabes « faibles »)

Au-delà de l'exemple des suffixes désaccentués du mandarin évoqués ci-dessus, des syllabes « inaccentuées » (ou syllabes « faibles ») peuvent apparaître dans diverses positions, selon les langues : en particulier les présyllabes de sesquisyllabes (quasi-disyllabes dont la première syllabe est réduite), et certains mots grammaticaux.

> In some languages we find syllables which never occur outside of larger units and never have a tone of their own. These syllables are pronounced with markedly less

> stress than the neighboring ones, and often the quality of their vowel is not distinctive. (Mazaudon 1977:79, qui cite Henderson 1961:60)

Le fait que les mots grammaticaux aient tendance à connaître une réalisation affaiblie est bien établi au plan synchronique dans de nombreuses langues ; ses conséquences diachroniques sont également connues : les changements phonétiques sur les mots grammaticaux ne sont pas nécessairement conformes aux régularités décelées sur les mots lexicaux ; cette différence peut avoir part à l'apparition d'une classe de syllabes sans ton, comme en chinois mandarin standard. Au plan de la description synchronique des langues à tons, la notion d'*extramétricité* rejoint celle d'*absence d'accent* ; cette dernière notion peut même permettre de faire l'économie de la notion d'extramétricalité, la notion d'accent apparaissant suffisante pour caractériser le comportement des particules « faibles » aussi bien que celui des syllabes « faibles » à l'intérieur de mots lexicaux. Il paraît également possible d'employer le terme d'*accent* pour caractériser les présyllabes réduites des langues, couramment appelées *sesquisyllabiques*, qui possèdent des quasi-disyllabes formées d'une présyllabe et d'une syllabe pleine. Les présyllabes sont prosodiquement faibles ; leur voyelle est centralisée (voir notamment Henderson 1952, Matisoff 1973b, Ferlus 1998b, Svantesson 2003:30). La syllabe principale serait « accentuée », au détriment de la présyllabe :

> Cambodian disyllables are of two patterns, viz : (a) A sequence of two major syllables (b) A minor syllable followed by a major syllable, with stress on the latter. (Henderson 1952:150)

De même dans la description du khmu (kammu) par Suwilai :

> The phonological word in Khmu consists of one, two or three syllables. The stress occurs on the main final syllable of the word. (Suwilai 2003:21)

L'emploi de la notion d'« accent » peut paraître peu économique, du fait que cet emploi amène à conclure que la majorité des syllabes de la langue sont accentuées, et seule une minorité non accentuée : la catégorie (a) évoquée par E. Henderson, celle des disyllabes formés de deux syllabes pleines, serait une catégorie de *mots dont chacune des deux syllabes est accentuée*. Pour autant, le phénomène présente bel et bien des similarités avec les alternances accentuelles d'une langue comme l'anglais ; la réduction vocalique que connaissent les présyllabes rappelle les voyelles réduites des syllabes inaccentuées en anglais.

De façon assez similaire, les particules finales d'énoncé (« particules de discours », dont des exemples sont fournis par les phrases-cadres décrites au chapitre *Intonation pragmatique*, sections 2.2.2 et 2.3.2) ont généralement un statut tonal à part : en chinois et en thai, elles n'ont pas de ton propre (Luksaneeyanawin 1998:390) ; en vietnamien, elles portent un ton lexical au même titre que les autres syllabes, mais le marquage des frontières (*accent démarcatif* de la terminologie évoquée en section 3.1.1) a lieu sur la dernière syllabe lexicale qui précède les particules finales (ce qui représente une difficulté pour l'apprenant français,

qui aura d'abord tendance à réaliser la frontière sur la dernière syllabe, comme c'est le cas dans sa langue maternelle) ; les particules sont donc « extramétriques », et peuvent être dites « inaccentuées », tout en précisant que cette absence d'accent ne commande pas de neutralisation tonale. En naxi, où les particules finales portent, comme en vietnamien, un ton lexical, certaines sont extramétriques, d'autres non ; de même en birman, où certaines particules sont exclues du compte des syllabes et de leur groupement en pieds dans la versification traditionnelle (Denise Bernot, communication personnelle ; quelques indications à ce sujet figurent dans Hla Pe et Bernot 1974, « La poésie birmane: 1300-1971 »).

Pour conclure, la notion d'accent a une utilité dans la description des langues à tons ; dans l'esprit des réflexions de Martine Mazaudon (Mazaudon 1977:79-81), elle peut être appliquée pour unifier la description des diverses « syllabes faibles », qu'elles soient initiales ou finales de mot, ou qu'il s'agisse de syllabes « extramétriques » au plan du découpage de l'énoncé en constituants prosodiques. La section qui suit est consacrée à un ensemble (relativement hétérogène) de systèmes prosodiques pour lesquels la question des relations entre tons et accent se pose avec une acuité particulière.

3.1.6.2. Les langues à accent mélodique forment-elles un pôle typologique distinct ?

Parmi les langues à accent mélodique les plus couramment citées figurent le suédois et le serbo-croate (notamment décrits par Lehiste 1970), le japonais de Tokyo, et le basque. Ce groupe inclut également des langues africaines telles que le luganda. **La présence ou l'absence d'un accent mélodique, ainsi que sa position dans le mot, sont spécifiées lexicalement.** L'accent mélodique du japonais de Tokyo est généralement analysé comme une séquence de tons ponctuels, HL : le ton H s'associe à la syllabe porteuse de l'accent mélodique, et le ton L à la syllabe suivante (ainsi qu'aux syllabes suivantes, en l'absence de frontières ou d'autres accents mélodiques). **Le nombre d'accents mélodiques varie selon les langues** : le japonais de Tokyo n'en connaît qu'un ; le barasana possède deux accents mélodiques, H et HL (Gomez-Imbert et Kenstowicz 2000) ; le dialecte japonais d'Osaka, deux accents mélodiques également : HL et LHL. Les systèmes à accent mélodique sont, selon les cas, plutôt proches de systèmes de tons, ou plutôt proches de systèmes d'accent. Certains dialectes japonais se rapprocheraient de systèmes à accent, notamment le parler des îles Ryukyu du sud :

> Here, a phonological rule applies to move the H-tone to a syllable not immediately adjacent to it, but to one at a greater distance. This type of long-distance shifting of the H-tone may be better comprehended with recourse to « accent » rather than « tone ». Such attraction of the prominent peak to the penultimate syllable, which is a metrically unmarked position for stress and a target of stress assignement for many languages, suggests the realization of the edge-marking function of accent in this dialect (Odden 1999) ; thus it might be possible to postulate a system starting to exhibit a (at least partially) stress-like characteristic in this case. (Matsumori 2001)

Les dialectes basques, géographiquement voisins de langues à accent, présentent une variation entre accent et accent mélodique (voir Hualde, Elordieta *et al.* 2002, « From pitch accent to stress accent in Basque »).

Dans certaines langues, la question d'une description en termes d'*accent mélodique* ou de *tons* se pose avec acuité. Parmi les langues qui se situent, pour reprendre le titre d'une conférence récente, « between stress and tone »[1], citons en premier lieu les langues bantu [bantoues], qui paraissent fournir un excellent exemple du continuum entre langues à tons ponctuels, plus conservatrices, et langues qui possèdent un système d'accent mélodique, aussi appelé « accent tonal », lequel représente une innovation (voir notamment Philippson 1998). Se rattachent également à cette problématique les systèmes prosodiques de langues de Nouvelle-Guinée, d'abord analysées comme ayant un ton syllabique, puis réanalysées comme ayant des accents mélodiques (Donohue 2003). Elsa Gomez-Imbert souligne que la famille tukano (langues d'Amazonie) comprend, outre des langues à accent mélodique comme le barasana, des langues à tons (Gomez-Imbert 2001).

Sur la base de ces observations, qui suggèrent que les systèmes à accent mélodique se trouveraient au milieu d'un continuum entre système tonal et système accentuel, **Larry Hyman conteste l'existence des systèmes à accents mélodiques comme pôle typologique distinct** :

> Although many scholars see the privative and often culminative nature of the underlying opposition to be evidence of an accentual system, I have here (and elsewhere) come down on the side of a tonal analysis of Japanese dialects. (Hyman 2001a:126)

Il esquisse une approche en termes de prototypes, que résument les propositions suivantes :

> 1) There are well-defined prototypes of tone- and stress systems, in which properties cluster as in the left- and right columns below.
> 2) There is no third type, i.e. no cluster of properties that defines all such languages as « pitch-accent » or « tonal-accent ».
> 3) So-called pitch-accent languages simply diverge from the two clear prototypes (tone and stress) by picking and choosing different properties from each column
> 4) Pitch-accent systems are thus different from each other, e.g. Tokyo Japanese≠Swedish≠Serbo-Croatian≠Bantu /H, Ø/, etc. (Hyman 2001b ; voir également la réponse de Kubozono 2001:264, qui défend l'existence des langues à accent mélodique comme pôle typologique).

---

[1] Cette conférence, qui s'est tenue du 16 au 18 juin 2005 à Leiden (Pays-Bas), avait pour argumentaire : « The Between Stress and Tone conference (BeST) focuses on the variety of word-prosodic systems in the languages of the world. Particular attention will be paid to systems that challenge the typological classification into word stress, lexical pitch-accent and tone. We also invite studies that discuss the interaction between the phonetic encoding of word-prosodic features and utterance-level prosody. »

Afin de prolonger cette réflexion, la section 3.1.6.2 revient sur une question soulevée par la combinaison des tons et de l'accent : peut-il exister, dans une même langue, des tons lexicaux et un accent lexical qui varient indépendamment ?

3.1.6.2. Une libre combinaison des tons et d'un accent lexical est-elle possible ? L'exemple du ma'ya

Certaines langues possèdent à la fois des tons et un accent lexical ; en revanche, à ce jour, il n'a pas été observé de langue combinant un ton par syllabe et un accent lexical distinctif. Remijsen 2002 dit apporter un contre-exemple à cette généralisation, mais le système prosodique en question, celui de la langue ma'ya, est réanalysé par Wetzels 2002 comme un système d'accent mélodique.

Il est bien établi qu'une langue à tons peut posséder des syllabes non accentuées, l'absence d'accent se traduisant par une neutralisation partielle ou totale des contrastes tonals (et éventuellement des contrastes vocaliques). Ce mode de combinaison entre tons et accent n'est pas *orthogonal* : tons et accent ne varient pas de façon indépendante, au sens où l'intégralité de la combinatoire entre tons et accents serait explorée.

Une étude récente défend néanmoins l'idée selon laquelle il existerait au moins une langue dont le système prosodique comporterait des tons lexicaux et un accent lexical qui varieraient indépendamment. **Bert Remijsen juge qu'en ma'ya, langue austronésienne** (malayo-polynésienne de l'Est)**, les tons lexicaux vont de pair avec un accent lexical distinctif** (Remijsen 2002, « Lexically contrastive stress accent and lexical tone in Ma'ya »)**.** Autrement dit, les contrastes tonals ne seraient pas neutralisés en l'absence d'accent.

Pourtant, de l'analyse de Bert Remijsen, il ressort que l'intégralité des combinaisons possibles entre tons et « accent » n'est pas attestée en ma'ya. En reportant les faits décrits par cet auteur dans un tableau, nous avons abouti au bilan suivant (tableau 4.4) :

ton *(lexical tone)*	**haut**	**montant**	**pas de ton**
Accent lexical *(lexically contrastive stress accent)*	ton haut *(high tone)*	ton montant *(rising tone)*	ton descendant *(falling tone)*
Pas d'accent lexical *(absence of lexically contrastive stress accent)*	ton haut *(high tone)*	ton montant *(rising tone)*	ton bas *(low tone)*

Tableau 4.4. Présentation des combinaisons entre tons et « accent » en ma'ya d'après l'analyse de Remijsen 2002

Le tableau 4.4 montre que l'« accent » postulé par Remijsen est sans effet sur les syllabes tonalisées : il n'affecte que les syllabes sans ton, lesquelles reçoivent un ton bas par défaut, un ton descendant sous l'accent lexical. Autrement dit, ce ne sont pas les distinctions tonales qui seraient neutralisées en l'absence d'accent : c'est la distinction entre présence et absence d'accent qui est neutralisée par la présence d'un ton lexical. En l'absence de ton lexical, l'accent est contrastif, et se manifeste par une mélodie.

Au vu de ces faits, Leo Wetzels suggère de relativiser la proposition formulée par Remijsen ; reprenant les faits exposés par ce dernier, il fait ressortir la similarité entre le ma'ya et le dialecte néerlandais de Limburg, qui possède à la fois un accent de mot et une opposition tonale binaire limitée aux syllabes accentuées qui contiennent deux mores sonantes. Une telle neutralisation des contrastes de tons en syllabe non accentuée est une situation bien connue (voir par exemple ci-dessus, section 3.1.6.1) ; au bilan, il apparaît que tons et accent ne varient pas de façon indépendante en may'a[1].

Dans l'attente d'éventuelles découvertes, il paraît possible de faire l'hypothèse qu'il n'existe pas de langues dans lesquelles distinctions tonales et distinctions accentuelles seraient indépendantes[2].

3.2. Une nouvelle perspective typologique, indépendante de la présence ou de l'absence de tons lexicaux : langues à *prosodie calculée* et langues à *prosodie contrôlée par l'énonciateur*

Le panorama proposé ci-dessus souligne la variété des systèmes prosodiques. **Le fait qu'une langue possède tons ponctuels, tons mélodiques, accents mélodiques ou accent de mot ne préjuge pas du détail de son système prosodique**. Larry Hyman formule les remarques suivantes :

> ... only certain parameters impress linguists enough to establish language types. Within phonology, there is a class of "click languages", but not "implosive languages", "open syllable languages" (...).The question is whether there is any reality in classifying languages rather than analyzing the properties of the relevant

[1] La description des systèmes prosodiques des dialectes rGyalrong de Caodeng et de Luoxi se pose dans des termes similaires, d'après ce que nous en a dit Jackson Sun (communication personnelle, et article en préparation).

[2] Parmi les systèmes prosodiques décrits comme possédant un accent qui se combinerait librement avec les tons, mais dont un réexamen suggèrent qu'ils ne comportent pas un tel accent lexical, figurent les langues yi, géographiquement voisines du naxi et ses proches parentes au plan du classement en familles de langues. D'après la description de Chen Shilin, Bian Shiming *et al.* 1985:21-24, celles-ci comporteraient, outre des tons lexicaux ponctuels et un contraste entre réalisation *tendue* et *non tendue* des tons, un accent lexical distinctif. (Dans les exemples qui suivent, l'apostrophe signale l'accent lexical.)

/vi˩ si˧/ : « hôte, personne qui invite »

/'vi˩ si˧/ terme générique désignant l'hôte et ses invités

Néanmoins, l'absence de notation de l'accent dans le lexique et les énoncés fournis dans l'ouvrage en question suggère que l'accent lexical a tout au plus un rôle marginal. Cet « accent » participe à des phénomènes expressifs (/ha˥ ko˧˦ / « là-bas », /'ha˥ ko˧˦ / « tout là-bas au loin » ; le symbole ˧˦ désigne un ton moyen accompagné d'une qualité de voix tendue) ainsi qu'à la démarcation des groupes intonatifs *(phrasing)*, ce qui suggère pareillement qu'il s'agit d'un phénomène intonatif plutôt que lexical. David Bradley (communication personnelle) indique ne jamais avoir rencontré d'accent lexical distinctif dans les langues yi (ni), dont il est spécialiste.

> subsystems. This issue is particularly pronounced in the area of prosody, where the practice of distinguishing "stress languages" from "tone languages" is well established and has encouraged some to propose additional types: "pitch-accent language", "restrictive tone language", and so forth. However, as Greenberg (1974:14) notes, "the same data can be utilized either for a typology of linguistic properties or a typology of individual languages." A major aim of this study is to show that there is considerably more diversity in prosodic systems than such labels have thus far distinguished. (Hyman 2005:1)

En typologie prosodique comme en typologie morphosyntaxique, certaines caractéristiques linguistiques sont corrélées, mais ces corrélations n'ont pas force de loi. Greenberg 1963 l'indique clairement dans ses formulations : par exemple « If in a language the verb follows both the nominal subject and nominal object as the dominant order, the language *almost always* has a case system » (c'est nous qui soulignons). Il existe en outre des caractéristiques linguistiques qui paraissent faiblement liées aux autres caractéristiques de la langue concernée : la différence entre structure actancielle *accusative* et *ergative* ne paraît pas avoir de conséquences étendues sur la morphosyntaxe de la langue dans son entier : Lazard 1994:34, qui distingue non pas deux mais cinq types de structures d'actance, souligne qu'il n'est pas rare que plusieurs d'entre elles se trouvent en usage dans la même langue. Boyd Michailovsky (séminaire Langues tibéto-birmanes) évoque avec humour la conception *mystique* de l'ergativité qu'auraient certains auteurs, dont les généralisations sur ce pôle typologique supposé sont battues en brèche par des langues tibéto-birmanes à ergatif telles que le hayu (Michailovsky 1988a) et le limbu (Van Driem 1987, Michailovsky à paraître).

Plutôt que la définition de catégories, la tâche de la typologie prosodique paraît être, en premier lieu, de proposer un ensemble de caractéristiques pertinentes pour décrire les systèmes rencontrés dans les langues[1]. Pour reprendre les termes de Larry Hyman (séminaire Langues tibéto-birmanes, 2005) :

> Un rêve...
> Si seulement on avait pour chaque groupe de langues tonales (le mixtec, le bantou, le kuki-chin etc.) un chapitre bref et concis contenant une présentation lucide des caractéristiques tonales principales pour guider nos enquêtes tonales et notre compréhension de la tonologie en général.

Les réflexions qui suivent visent à aller dans cette direction en se fondant sur les observations publiées sur les langues à tons ponctuels d'Afrique et d'Asie pour faire ressortir une caractéristique qui paraît pertinente dans la caractérisation des systèmes prosodiques : le

[1] Cette réflexion ne revient bien sûr pas à abandonner entièrement l'ambition de déceler des pôles typologiques caractérisés par l'association privilégiée de plusieurs caractéristiques structurelles (dans l'esprit des recherches de Bernard Comrie [Comrie 1981, en particulier pp. 35-37] et William Croft [Croft 1990]), mais vise à souligner que l'entreprise doit être conduite avec prudence.

degré de liberté intonative laissé à l'énonciateur dans la réalisation des syllabes successives de l'énoncé.

3.2.1. Intuition de départ

Un aperçu typologique suggère que la prosodie de langues tonales et non tonales peut présenter de grandes similarités, la présence de tons lexicaux en elle-même limitant assez peu la marge de variation intonative. (Ce constat sera affiné à mesure de la discussion : la présence de plus de trois tons *ponctuels* paraît limiter plus fortement la variation intonative que la présence de tons-contour « de type asiatique » [comme ceux du vietnamien], quel que soit le nombre de ces derniers.) Une différence plus importante paraît résider dans la latitude de jeu prosodique que la langue laisse à l'énonciateur. Nous proposons de distinguer

- langues à prosodie *contrôlée par l'énonciateur*, dans lesquelles ce dernier est à tout instant libre de jouer sur divers paramètres prosodiques pour façonner son discours

et

- langues à *prosodie calculée*, qui ne présentent pas cette latitude de jeu.

Les trois langues qui ont fait ici l'objet d'une étude expérimentale apparaissent plutôt comme des langues à prosodie contrôlée : le naxi (langue à tons ponctuels), le vietnamien (langue à tons modulés complexes), et l'anglais (langue à accent). Pour affiner, parmi les trois, c'est le naxi qui est le moins proche du prototype de la *langue à prosodie contrôlée*, du fait de la présence dans cette langue de quelques processus de *calcul tonal* (ajout d'un ton H à une syllabe au ton L ou M à des fins d'intensification, ton H flottant se déplaçant vers la syllabe située à sa gauche), tandis qu'anglais et vietnamien seraient exemplaires du type *prosodie contrôlée*.

Anglais, vietnamien et naxi possèdent également des moyens grammaticaux de mise en valeur. En naxi, ces moyens sont brièvement décrits au chapitre II, section 1.1. L'anglais (de même que le chinois et le français) possède, entre autres moyens de mise en valeur, des constructions clivées (pour le chinois, voir en particulier Paris 1979, 1998, 1999). En français, les structures clivées (« c'est... qui... ») sont traditionnellement citées comme équivalent de la mise en valeur intonative de l'anglais[1] ; néanmoins, leur fréquence dans les corpus de français oral est basse, indice indirect de l'importance des moyens intonatifs de

[1] « En français, l'accent d'insistance n'a pas pour effet d'attirer spécialement l'attention de l'auditeur sur un concept particulier, mais d'exprimer un certain état d'esprit du locuteur. Si, dans cette langue, on veut mettre en valeur un des mots d'un énoncé, on aura recours à une modification de la construction et à des morphèmes particuliers, non à un accroissement de l'intensité avec laquelle est articulée une des syllabes du mot. L'Anglais qui veut faire savoir que c'est Pierre qui vient, et non Paul ou tel autre, prononcera, dans *Peter comes*, la syllabe initiale de *Peter* (celle qui porte l'accent du mot) avec une intensité particulière. (...) En français, on dira dans ce cas *c'est Pierre qui vient*. On est tenté d'expliquer les réactions particulières du français sur ce point en énonçant la règle que seules les langues qui connaissent un net accent de mot sont susceptibles de le renforcer encore pour mettre en valeur tel ou tel mot de la phrase. » (Martinet 1956:100)

focalisation[1]. **La présence de structures clivées dans une langue n'est nullement exclusive du marquage intonatif de la hiérarchie de l'information.**

A ce jour, les langues à prosodie calculée les plus typiques semblent être exclusivement des langues africaines (Clements 1979, 1984 ; Rialland et Sangaré 1989). Annie Rialland souligne le rapport direct, quasi bijectif, entre les processus phonologiques catégoriels de calcul des séquences tonales et les réalisations phonétiques dans plusieurs de ces langues (Rialland 2001, « Anticipatory Raising in Downstep Realization: Evidence for Preplanning in Tone Production » ; voir également Laniran 1992:273). La distinction de hauteur entre tons phonologiquement identiques y est essentiellement réservée au phénomène (non distinctif) de *downstep automatique* : dans les séquences HLHLHL, les H se succèdent à une hauteur chaque fois diminuée. Ce phénomène régulier est asymptotique : les trois premiers abaissements du ton H sont nettement visibles sur la courbe mélodique ; les suivants sont atténués, voire absents. En dagara, langue Niger-Congo (branche gur) parlée au Burkina Faso, les séquences de tons H sont également réalisées par une courbe mélodique plane[2]. En revanche, les abaissements de tons H successifs sont distinctifs (phénomène de *downstep non automatique*) ; ils sont déclenchés par des tons bas flottants (Rialland 2001:303). Le *downstep* n'est pas asymptotique comme en yorùbá : les locuteurs seraient en mesure, dans une élocution soignée, de réaliser phonétiquement les *downsteps non automatiques* tout au long de l'énoncé, recourant à des stratégies d'anticipation (variables selon les locuteurs). L'interprétation paraît simple : si, comme en yorùbá, l'abaissement était asymptotique, les tons H seraient uniformément rabaissés après les trois ou quatre premières manifestations du *downstep*, ce qui compromettrait la bonne réalisation du *downstep non automatique*, lequel est porteur d'information linguistique[3].

---

[1] Voir Delais-Roussarie, Rialland *et al.* 2002 et références citées, et la réflexion générale de Paul Cappeau sur l'absence dans les corpus d'oral de certaines constructions attendues (Cappeau 2005).

[2] Les courbes montrent en réalité une légère pente descendante, qui varie entre locuteurs ; la comparaison entre données yorùbá et données naxi proposée au chapitre II (1e partie, section 3) suggère que la baisse de fréquence fondamentale dans les énoncés déclaratifs (prononcés en élocution soignée) dont toutes les syllabes portent un ton H est moindre en yorùbá qu'en naxi.

[3] Les études au travers desquelles ces architectures tonales complexes sont connues recourent pour l'essentiel à de la parole lue ; beaucoup des phrases employées prêtent en outre le flanc à la critique d'artificialité (Ivan Fónagy remarque de façon générale que les exemples rencontrés dans les études linguistiques sont « trop souvent insipides ou grotesques » [Fónagy 1989]) ; de plus, les études instrumentales demeurent minoritaires dans les descriptions des systèmes prosodiques africains. Ces trois limites de la documentation existante sur les langues africaines tendent à conforter dans leur incrédulité certains chercheurs non spécialistes de langues africaines, selon qui les propriétés linguistiques attribuées aux langues concernées reflètent les *a priori* des descripteurs plutôt que les faits linguistiques eux-mêmes. Il paraît donc utile de rappeler que la variabilité bien connue des faits de parole spontanée n'invalide pas les conclusions établies au sujet de la parole lue : ainsi, l'emploi de parole lue a permis des progrès considérables dans la compréhension de l'intonation du français

3.2.2. Formulation proposée par Larry Hyman : tons paradigmatiques et tons syntagmatiques

Larry Hyman (communication personnelle) propose d'opposer « tons paradigmatiques », qui offrent une latitude de variation (catégorie rejoignant l'ensemble « langues à prosodie contrôlée par l'énonciateur »), et « tons syntagmatiques » dont la réalisation se déduit de la séquence tonale dans laquelle ils apparaissent (rejoignant le pôle des « langues à prosodie calculée »). Ces deux pôles sont séparés par une échelle de degrés ; ce qui détermine la position d'une langue donnée sur cette échelle est l'étendue du paradigme tonal ouvert en tout point de l'énoncé. Une langue qui possède trois tons (haut, moyen et bas) et un *downstep non automatique* (abaissement d'un ton dû à la présence d'un ton flottant qui le précède) présente cinq possibilités après un ton H, comme l'illustre le cas du ngamambo (tableau 4.5).

	notation	description
1	H	ton haut
2	M	ton moyen
3	L	ton bas
4	!L	ton bas subissant un *downstep* du fait d'un ton flottant
5	L°	ton bas qui est réalisé sans descente phonétique (à la différence de L) du fait d'un ton flottant

Tableau 4.5. Paradigme tonal ouvert en ngamambo après une syllabe au ton H

Dans ces conditions, la marge de variation qui demeure est fortement réduite. **Dans les langues « très syntagmatiques » comme le ngamambo, la courbe mélodique de surface pourrait se déduire de la séquence de tons qui est obtenue par application des règles, souvent complexes, qui déterminent la possibilité ou non de certains contours dans certaines positions syntaxiques, et les domaines à l'intérieur desquels les tons se déploient et interagissent.** (Des modèles équivalents peuvent être proposés, dans une autre perspective théorique, en termes de *sélection*, en plusieurs étapes, de la séquence de tons finale ; cette perspective très débattue actuellement, celle de la théorie de l'optimalité, n'est pas développée ici.) Les frontières intonatives sont marquées par des tons, dont la réalisation, au terme de l'application des diverses règles d'association entre syllabes et tons (tons lexicaux le cas échéant, et tons de frontière), ne se distingue pas de celle des tons lexicaux.

> Les règles d'insertion des démarcatifs s'appuient sur une information syntaxique qui fournit ce qui est nécessaire et suffisant pour prédire correctement les alternances phonologiques. Après avoir inséré les démarcatifs, l'information syntaxique disparaît pour céder la place à une représentation strictement linéaire des phrases. (Creissels et Grégoire 1993:132)

(Vaissière 1993, 1997) ; le fait d'écarter certains paramètres peut parfois être une abstraction utile, même si les paramètres en question ont incontestablement un rôle important dans la communication.

Des expériences en synthèse de la parole permettraient d'approfondir la connaissance expérimentale de ces systèmes prosodiques, qui paraissent se prêter de façon exemplaire à une synthèse automatique des courbes de fréquence fondamentale. De telles études conduiraient peut-être à nuancer l'idée selon laquelle les tons de frontière viennent se placer strictement sur le même plan (plus précisément sur la même *droite*) que les tons lexicaux. Annie Rialland (communication personnelle) souligne que, d'une langue à l'autre, il est possible d'hésiter sur l'interprétation de l'abaissement final d'énoncé : indice d'un ton bas final, ou phénomène distinct de la chaîne tonale. Néanmoins, dans l'ensemble, il paraît établi que dans les langues en question, la réalisation phonétique finale ne présente pas d'écarts notables par rapport à la séquence tonale calculée.

Par opposition à ce type *syntagmatique*, le type *paradigmatique* serait celui des « langues à déclinaison », qui typiquement ne connaissent pas d'interactions catégorielles entre tons telles que le *downstep* (entendu au sens de : *downstep* non automatique ; voir l'entrée correspondante du Glossaire). Les phénomènes de déclinaison ne peuvent être modélisés en termes de tons ajoutés localement : ils s'échelonnent du local au global, et connaissent des degrés. **Dans les langues à tons paradigmatiques, la courbe mélodique finale n'est pas prévisible d'après la séquence des tons en fin de *dérivation tonale*, celle-ci étant par ailleurs inexistante ou quasi-inexistante**, à la différence du pôle typologique précédemment décrit. **Les *contours* peuvent être porteurs d'information sans être catégoriels**. (Telle est la situation observée en naxi ; cf. ci-dessus, chapitre II, *passim*.) **La réalisation des tons est sensible à la charge informationnelle que le locuteur leur confère (*prominence-sensitive realisation of tone*).**

Cette perspective typologique souligne que l'intonation est plus fortement contrainte par la présence des phénomènes de *calculs sur les séquences tonales* (d'où l'idée de l'expression « langues à tons syntagmatiques » proposée par Larry Hyman, équivalent de notre expression « langues à prosodie calculée ») **que par la présence des tons lexicaux en elle-même.** Le vietnamien (« langue à tons paradigmatiques », ou « langue à tons contrôlés par l'énonciateur ») fournit l'exemple d'une langue au paradigme tonal complexe qui connaît néanmoins des phénomènes de mise en valeur intonative locale.

Le degré de *calcul des séquences tonales*, ou inversement *de jeu énonciatif sur la réalisation des tons*, a des conséquences au plan phonétique. Dans les langues où le calcul tonal est central dans la prosodie, dans une séquence de tons identiques, une cible basse (ou une cible haute, lorsqu'elle est précédée d'une cible basse) tend à être atteinte tardivement. Pour reprendre l'exemple de la forme expressive bambara [bòròbárá] (mentionnée au chapitre II, 1er volet, section 2.1.3), qui est porteuse d'un schéma L + H, chacun des tons se déployant sur une des moitiés de la forme réduplicative (Dumestre 1987:95) :

bo ro ba ra



L H

Phonétiquement, la cible, haute ou basse, est atteinte tardivement : sur la seconde syllabe, de sorte que le schéma HH se réalise par une légère montée (notation informelle : H^H), le schéma LL par une descente (notation informelle : L_L). A l'inverse, en naxi, la hauteur relative de tons H qui se succèdent est déterminée par le statut informationnel des mots concernées (voir chapitre II, 2e volet), et par l'éventuelle présence de frontières (voir chapitre II, 1er volet).

3.2.3. Le sandhi *est plutôt une caractéristique d'une langue à prosodie contrôlée, le* downstep *plutôt une caractéristique d'une langue à prosodie calculée*

Il serait illusoire de prétendre réduire la variété des systèmes typologiques aux deux types esquissés ici, qui représentent les deux extrêmes d'un continuum. Avec cette réserve, il paraît néanmoins possible de prolonger cette perspective typologique en soulignant les affinités de l'un et l'autre pôle typologique avec certains phénomènes de changement tonal. Le *downstep* serait un processus relativement fréquent dans les langues présentant un degré élevé de *calcul prosodique*, tandis que le *sandhi tonal* serait plus fréquent dans les langues où la variation allophonique des tons est en grande partie sous le contrôle de l'énonciateur.

Un élément de contraste entre sandhi et *downstep* est que le caractère catégoriel du sandhi est parfois problématique, tandis que le *downstep* apparaît comme un excellent exemple de phénomène catégoriel. Le sandhi tonal par lequel une séquence *ton 3 + ton 3* devient *ton 2 + ton 3* en chinois mandarin attire depuis longtemps l'attention des phonéticiens et phonologues ; il tend à faire oublier les autres cas de sandhi (déjà mentionnés dans les descriptions de Chao Yuen-ren). Beaucoup de ces derniers n'ont pas de statut catégoriel établi, et ne représentent peut-être que le figement progressif d'habitudes co-articulatoires. La coarticulation tonale (variation du ton en fonction du ton des syllabes environnantes) est de grande ampleur en mandarin (Xu Yi 1997, « Contextual tonal variations in Mandarin »), de même qu'en vietnamien (Han et Kim 1974:223, « Phonetic variation of Vietnamese tones in disyllabic utterances ») ; au plan phonétique, la variation allophonique des tons en contexte est considérable ; le sandhi a sans doute partie liée avec ces phénomènes, comme le note Matthew Chen *(Tone Sandhi : Patterns Across Chinese Dialects)* :

> It is not clear whether it is desirable or even possible to segregate tonal coarticulation from tone sandhi proper. (Chen 2000:25)

3.2.4. Choix terminologique

Un choix terminologique demeure ouvert pour désigner les deux pôles typologiques caractérisés ci-dessus :

premier pôle typologique	second pôle typologique
langue à tons syntagmatiques	langue à tons paradigmatiques
langue à prosodie calculée	langue à prosodie contrôlée par l'énonciateur

La première désignation (tons syntagmatiques et tons paradigmatiques), proposée par Larry Hyman, recourt à des termes qui font partie du *vocabulaire de base* de la linguistique, et sont par là porteurs de connotations qui peuvent être indésirables : *tons syntagmatiques* évoque plus particulièrement l'intonation syntaxique (ou *phrasing*), et les tons de frontière présents dans certaines langues africaines, ce qui n'est pas souhaitable s'agissant d'un terme qui vise à caractériser un système prosodique tout entier. Pareillement, *tons paradigmatiques* a l'inconvénient de s'appliquer à tous les tons lexicaux : les tons des langues du premier pôle typologique (« tons syntagmatiques ») composent pareillement des paradigmes, ce qui en un sens est la définition de tout ton ou schème tonal lexical. En outre, l'expression demande précision au sujet du paradigme en question : s'agit-il du paradigme lexical (par exemple, le paradigme de six tons des syllabes à finale sonante du vietnamien), ou d'un paradigme de variantes intonatives d'un même ton lexical, entre lesquelles l'énonciateur est libre de choisir ? *Last but not least*, le choix d'inclure la notion de ton dans ces désignations limite d'emblée la portée typologique de l'observation au domaine des langues à tons ; or la présente étude expérimentale montre que la présence ou l'absence de tons lexicaux n'a pas en elle-même une influence déterminante sur le mode de structuration prosodique. L'expérience fait en effet ressortir des similarités entre l'anglais, le naxi et le vietnamien, langues qui toutes trois sont plutôt proches du second pôle typologique (« langues à prosodie contrôlée par l'énonciateur ») ; inversement, les langues à *prosodie calculée* ne sont pas nécessairement de langues à tons lexicaux : le wolof paraît fournir l'exemple d'une langue sans tons mais à *prosodie calculée* (Rialland et Robert 2001, 2003). Les termes de *prosodie calculée* et *prosodie contrôlée* ont l'avantage de s'appliquer aux langues les plus diverses, qu'elles aient ou non des tons lexicaux.

Bilan

Les spéculations qui précèdent, dont la généralité est certes imprudente, s'autorisent du projet de jeter les bases d'une typologie, au-delà du constat de la considérable variété des combinaisons entre traits prosodiques observées dans les langues.

L'inclusion d'une distinction entre *prosodie calculée* et *prosodie contrôlée par l'énonciateur* dans les descriptions prosodiques peut paraître de peu d'utilité dans la mesure où la quasi-totalité des langues prises en compte dans les discussions prosodiques actuelles sont proches du second de ces pôles typologiques. Elle a pourtant une conséquence importante : dépasser la querelle théorique entre *modélisation de la prosodie en une séquence de tons* et *modélisation de la prosodie en termes d'interaction de facteurs de plusieurs niveaux*, en suggérant que ces deux approches ont un degré de pertinence variable selon les langues.

Table des matières de la Conclusion

1. Conclusions au plan de la technique expérimentale ____ 225

1.1. Possibilité d'une évaluation du quotient ouvert par électroglottographie, même dans les cas de phonation non modale ____ 225

1.2. Le quotient ouvert n'est pas déductible de F_0 ____ 225

1.3. Utilité de la mesure d'intensité acoustique globale dans la comparaison des mêmes mots dans différents contextes d'énonciation ____ 226

2. Conclusions concernant la prosodie des trois langues étudiées ____ 226

2.1. Conclusions au sujet du vietnamien : données nouvelles sur la qualité de voix associée aux tons, et perspectives typologiques ____ 226

2.2. Conclusions au sujet du naxi ____ 227

Le naxi n'utilise pas la qualité de voix de façon lexicalement distinctive ____ 227

Tons flottants, réduction syllabique et contours intonatifs dans la prosodie naxi ____ 227

2.3. Conclusions au sujet de l'anglais ____ 228

2.4. Utilité de la comparaison entre langues ____ 228

3. Conclusions au plan de la modélisation prosodique et de la typologie ____ 229

3.1. Confirmation de l'hypothèse d'un partage des ressources entre accentuation et intonation ____ 229

3.2. Variété typologique. Proposition de deux pôles typologiques ____ 229

3.2.1. Réflexions au sujet des modèles autosegmentaux-métriques de l'intonation ____ 230

3.2.2. Une perspective typologique nouvelle : langues à prosodie calculée et langues à prosodie contrôlée par l'énonciateur : différences typologiques dans le degré de contrôle de la prosodie par l'énonciateur ____ 230

Conclusion

Après un bilan au plan de la technique expérimentale seront présentées les conclusions auxquelles l'étude est parvenue

- au sujet de la prosodie des trois langues concernées
- au plan des modèles théoriques et de la typologie.

1. Conclusions au plan de la technique expérimentale

1.1. Possibilité d'une évaluation du quotient ouvert par électroglottographie, même dans les cas de phonation non modale

L'étude confirme que le calcul du quotient ouvert par la détection des pics positifs et négatifs sur la dérivée du signal électroglottographique fournit des indications précises sur la qualité de voix (conformément aux conclusions de Henrich, d'Alessandro *et al.* 2004), même dans les cas de phonation non modale.

1.2. Le quotient ouvert n'est pas déductible de F_0

Une augmentation de quotient ouvert lors d'une augmentation de F_0 est observée par Holmberg, Hillman *et al.* 1989 (« Glottal air flow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch »), et Hanson, Gerratt *et al.* 1990 (« Frequency, intensity and target matching effects on photoglottographic measures of open quotient and speed quotient »). Néanmoins, le quotient ouvert n'est pas déductible de F_0. Une légère tendance à la corrélation entre F_0 et quotient ouvert ressort des données dans le cas du mécanisme phonatoire 1 (« voix de poitrine » des locuteurs masculins ; au sujet des mécanismes phonatoires, voir l'entrée correspondante du Glossaire) ; pour autant, il ne s'agit que d'une tendance. Ainsi, le locuteur naxi M5 présente l'étagement inverse : le quotient ouvert du ton haut est plus bas que celui du ton moyen, lui-même plus bas que celui du ton bas. En revanche, la locutrice F2 présente, dans les mêmes conditions expérimentales, un quotient ouvert plus élevé au ton haut qu'au ton moyen, plus élevé au ton moyen qu'au ton bas. Chez le locuteur M9, les valeurs sont équivalentes aux trois tons. En vietnamien, les valeurs de quotient ouvert des tons B2 et D2 s'opposent de façon systématique (valeurs très basses au ton B2, relativement élevées au ton D2), tandis que la plage de F_0 des deux tons est comparable.

La qualité de voix varie donc de façon relativement indépendante de F_0. Cette latitude est mise au service d'oppositions lexicales (en vietnamien), et de stratégies individuelles de mise en valeur locale (ce qui serait désigné, dans la tradition anglaise, comme *réalisation du nucleus*), cela dans les trois langues étudiées : en anglais, en naxi, et également en

vietnamien, où la qualité de voix associée à un ton est réalisée de façon plus appuyée par certains locuteurs lorsqu'une insistance particulière porte sur la syllabe concernée.

1.3. Utilité de la mesure d'intensité acoustique globale dans la comparaison des mêmes mots dans différents contextes d'énonciation

La mesure d'intensité acoustique globale est assez peu utilisée dans les études prosodiques, du fait de difficultés techniques, et de doutes sur sa pertinence linguistique. Les résultats obtenus suggèrent néanmoins que l'intensité acoustique est une mesure utile dans la comparaison des mêmes mots dans différents contextes d'énonciation, et qu'elle fournit une indication sur le degré d'*effort vocal*.

2. Conclusions concernant la prosodie des trois langues étudiées

Il convient de rappeler que nos expériences n'ont pas débouché sur une application en synthèse de courbes de F_0, étape décisive dans la modélisation. L'étude des données de production et de perception établit néanmoins un certain nombre de faits ; les principales conclusions au sujet de chacune des trois langues sont présentées ci-dessous.

2.1. Conclusions au sujet du vietnamien : données nouvelles sur la qualité de voix associée aux tons, et perspectives typologiques

Un point controversé de phonétique vietnamienne est tranché par l'étude expérimentale : les tons D1 et D2, qui apparaissent sur les syllabes à consonne occlusive finale (et sont en cela équivalents des ru[4] sheng[1] 入声 du chinois), ne comportent pas de glottalisation. Ce constat débouche sur la reconnaissance d'un inventaire typologique élargi des consonnes occlusives finales non relâchées : dans les langues décrites jusque-là, une glottalisation finale occasionnelle ou fréquente accompagne ces consonnes finales (dans le dialecte chinois du Fujian [Fukien] : Iwata, Sawashima *et al.* 1979 ; en chinois cantonais : Iwata, Sawashima *et al.* 1981 ; en thai : Harris 2001). L'établissement des faits vietnamiens conduit à rejeter l'hypothèse sur laquelle Graham Thurgood (Thurgood 2002) faisait reposer une réanalyse du modèle diachronique proposé par André-Georges Haudricourt (Haudricourt 1954, 1961)[1]. L'hypothèse que nous proposons pour expliquer la situation observée en vietnamien est que le ton D2 s'opposerait à un autre ton, le ton B2, en dépit du fait que l'un et l'autre ne

[1] Synchronie et diachronie ne s'opposent pas nécessairement dans la pratique du linguiste : « Il va sans dire que la pratique assidue d'une description synchronique dynamique prépare bien à faire, en diachronie, les hypothèses les mieux fondées au départ et les plus susceptibles d'être vérifiées. Il est, d'autre part, inutile d'insister sur l'avantage qu'il y a, dans la pratique diachronique, à opérer avec des synchronies dynamiques aussi rapprochées les unes des autres. (...) L'expérience montre qu'on évite difficilement de toucher, çà et là, à la diachronie dès qu'on cherche à donner, à une description synchronique, quelque profondeur. » (Martinet 1975:10)

puissent apparaître sur les mêmes syllabes. La réalisation des consonnes finales a par là partie liée avec les tons.

Au plan du vocabulaire, l'étude de la qualité de voix en vietnamien suggère qu'il est indiqué d'opposer deux phénomènes : la laryngalisation, qualité de voix plutôt détendue, qui peut être maintenue ; et la constriction glottale, qualité de voix tendue qui a typiquement pour effet d'interrompre la vibration des plis vocaux. Ces deux phénomènes sont liés, dans la mesure où une constriction glottale réalisée avec moins de force peut devenir une laryngalisation[1].

2.2. Conclusions au sujet du naxi

Le naxi n'utilise pas la qualité de voix de façon lexicalement distinctive

Les mesures de formants et de quotient ouvert confirment qu'à la différence des langues yi, ses voisines et relativement proches parentes, le naxi ne possède pas d'opposition entre voyelles tendues et voyelles relâchées, et que /ɯ/ et /ɚ/ sont deux voyelles distinctes en naxi ; faits bien connus (Fu Maoji 1940 ; Li Lincan, Zhang Kun *et al.* 1953 ; Rock 1963-1972 ; Mazaudon et Michailovsky 1979 ; Fu Maoji 1984 ; Fang Guoyu et He Zhiwu 1995 ; Pinson 1996, 1998), et répétés par Dai Qingxia (戴庆厦) en 1993[2], mais qui avaient été remis en cause par Yang Huandian (杨焕典) en 1984, et à nouveau en 1994[3]. Ce constat confirme une différence synchronique importante entre le naxi et les langues yi.

Tons flottants, réduction syllabique et contours intonatifs dans la prosodie naxi

Les observations expérimentales montrent une asymétrie de l'espace tonal : le ton haut possède une marge de variation intonative plus étendue que les deux autres tons de base. Or la modification locale des tons à des fins de mise en valeur intonative joue en naxi un rôle comparable à celui qu'elle possède en anglais. Le naxi connaît un processus catégoriel qui permet, dans certains cas, de contourner cette difficulté : l'ajout d'un ton haut (qui véhicule une *intensification*).

A également été mis en lumière (au chapitre II) le rôle du *jeu sur la déclinaison*. Dans le scénario diachronique esquissé pour expliquer les schémas de réduplication du naxi, le jeu

---

[1] Martinet 1975:239 formule le constat de la façon suivante : « ... la laryngalisation peut résulter d'un relâchement de la glottalisation ». A la différence de Martinet, nous employons « glottalisation » comme un terme générique désignant l'un et l'autre phénomène : laryngalisation et constriction glottale, suivant en cela Henton, Ladefoged *et al.* 1992:73.

[2] Article de Dai Qingxia intitulé « Le problème des voyelles tendues et relâchées du naxi », paru dans la revue chinoise *Minzu Yuwen* : "关于纳西语的松紧元音问题" , 民族语文1993年1期, pp. 27-31 et 36.

[3] Communication de Yang Huandian à la 27e Conférence internationale sur les langues et la linguistique sino-tibétaine (Paris, 1994) : « Nouvelles réflexions sur les voyelles tendues et relâchées en naxi – en réponse au professeur Dai Qingxia » (再论关于纳西语中的紧松元音问题 – 兼答戴庆厦教授).

sur la déclinaison aurait connu une phonologisation à un moment historique donné : au terme d'un épisode où le système prosodique se serait orienté vers l'instauration d'un *ton de mot*. (Ce scénario demeure au stade de l'hypothèse.)

2.3. Conclusions au sujet de l'anglais

L'étude établit que la qualité de voix varie selon des stratégies opposées d'un locuteur à un autre.

Au plan de la fréquence fondamentale, les changements observés lors du passage d'une condition de lecture à l'autre ne nous paraissent pas pouvoir être ramenés à une modification scalaire de tons catégoriels, comme le proposaient Pierrehumbert et Liberman 1984 : selon ces auteurs, lors d'une modification du *« degree of overall emphasis or excitement »*, le point terminal de la courbe correspondant au *nucleus* (valeur relativement basse, considérée par les auteurs comme la réalisation d'un ton L) resterait presque constant, tandis que le point le plus élevé de la courbe, modélisé comme un ton H, atteindrait une hauteur proportionnelle au degré d'insistance dont le mot est porteur. Les résultats conduisent à penser qu'une grande prudence est nécessaire dans l'application à l'anglais de concepts tels que « tons » et « accents mélodiques », cela dans le même souci de clarté terminologique qui sous-tend la remarque de Rialland 1995:367 au sujet de l'usage du terme « accent mélodique » (« pitch accent ») dans l'ouvrage *Autosegmental Studies on Pitch Accent* (van der Hulst et Smith 1988) :

> ...the book includes several chapters devoted to clearly non-pitch accent languages such as English, Dutch, and Hungarian, which do not satisfy the minimal condition generally assumed for a language to be a pitch accent system: that pitch must be lexically contrastive in some areas of the lexicon, at least at the level of surface realization. This condition is fulfilled in Japanese, Luganda, Swedish, and Serbo-Croatian, but not in the other languages just cited. At the very least, then, the title of the book is somewhat misleading, since it deals with the autosegmental function of tone in prosodic systems of almost all types (excluding only the ones situated at the "tonal pole" or very near it)—not just pitch accent systems in the usual sense.

Que l'anglais soit à la fois langue à accent mélodique et langue à tons (Pierrehumbert et Liberman 1984 emploient l'expression de « English tonology » pour décrire l'étude de l'intonation anglaise), voilà qui est beaucoup pour une langue... à accent (même remarque chez Fónagy 1989:75).

2.4. Utilité de la comparaison entre langues

C'est la comparaison entre langues, avec les mêmes outils, qui permet de faire ressortir la précision du contrôle de F_0 en naxi, et de la qualité de voix en vietnamien, mettant inversement en lumière la concomitance des variations de hauteur, d'intensité et de longueur en anglais.

Au-delà de la présente étude, les données quantifiées qui s'accumulent peu à peu permettent d'évaluer avec une précision croissante la place qu'occupe une langue donnée dans l'espace typologique, et d'effectuer des prédictions affinées sur l'emploi de F_0 et de la qualité de voix. Les résultats sur le naxi, le vietnamien et l'anglais (trois langues qui se rattachent aux pôles typologiques que sont, respectivement, les langues à tons ponctuels, les langues à tons complexes comprenant hauteur et qualité de voix, et les langues à accent) représentent autant de pierres de touche auxquelles rapporter les résultats obtenus pour d'autres langues[1].

3. Conclusions au plan de la modélisation prosodique et de la typologie

3.1. Confirmation de l'hypothèse d'un partage des ressources entre accentuation et intonation

La comparaison des données des trois langues montre que, parmi les corrélats de l'insistance (étudiés par le contraste de deux conditions de lecture), l'intensité acoustique tient une place relativement plus importante en naxi (le rapport entre variation d'intensité et variation de F_0 est plus bas qu'en anglais). **L'étude confirme et précise l'idée d'un partage des ressources entre accentuation lexicale (tons, accents mélodiques, ou accent) et intonation : les phénomènes lexicaux fournissent le cadre dans lequel la variation intonative prend son sens**.

Au plan de la modélisation prosodique, les résultats nous paraissent plaider pour une approche en termes d'interaction de plusieurs ordres de phénomènes, ce qui rejoint les modèles *superpositionnels* de l'intonation[2].

3.2. Variété typologique. Proposition de deux pôles typologiques

L'étude illustre la difficulté de la comparaison de l'intonation de langues éloignées ; toutefois, les obstacles à cette comparaison (tels que l'interaction entre intonation et morphosyntaxe) fournissent l'occasion d'une réflexion typologique.

---

[1] Seule une petite partie du corpus d'enregistrements audio et électroglottographiques rassemblé en collaboration avec d'autres chercheurs est utilisé dans la présente étude ; cet ensemble documentaire inclut désormais (outre trois dialectes naxi, le vietnamien et l'anglais) français, allemand, tcherkesse, birman et tamang. Dans cette quête expérimentale, il ne paraît pas irréaliste d'espérer qu'élargissement documentaire et affinement typologique continuent à progresser de conserve.

[2] Soulignons que la pratique didactique est éminemment compatible avec le cadre superpositionnel ; les idées superpositionnelles paraissent se prêter sans difficulté aucune à une application didactique. Le fait que l'enseignement soit, en un sens, en avance sur la théorie représente autant un encouragement dans la recherche qu'un gage de succès dans l'application de celle-ci.

3.2.1. Réflexions au sujet des modèles autosegmentaux-métriques de l'intonation

Rappelons la version extrême des modèles que Robert Ladd décrit comme *tone sequence models* : « ...the restrictive linear view says that all languages have tonal strings ; the main difference between languages with and without lexical tone is simply a matter of where the tonal specifications come from » (Ladd 1992). Les données expérimentales nous paraissent suggérer que la recherche de tons binaires sous-jacents à l'intonation revient à prêter un caractère *catégoriel* à des phénomènes dont il n'est pas évident qu'ils possèdent cette qualité, l'intonation étant dans une large mesure soustraite à la *double articulation* (Martinet 1960:83). Un modèle simplifié dans lequel l'attention se concentre sur la recherche de tons binaires ne paraît pas fournir les outils nécessaires pour décrire les phénomènes prosodiques observés en anglais, en naxi et en vietnamien[1]. Les catégories qui ont fait leurs preuves dans l'étude de nombreuses langues à tons d'Afrique (et également d'Asie) ne sont pas pour autant universelles ; certains traits prosodiques qui jouent un rôle central dans telle langue sont absents d'une autre[2]. La redécouverte progressive de l'importance des paramètres autres que F_0 pourrait amener les modèles intonatifs dits « autosegmentaux-métriques » à préciser peu à peu leur emploi du mot de « ton », qui gagnerait sans doute à être divisé en plusieurs notions, de façon à mieux refléter la diversité des phénomènes concernés.

3.2.2. Une perspective typologique nouvelle : langues à prosodie calculée *et* langues à prosodie contrôlée par l'énonciateur *: différences typologiques dans le degré de contrôle de la prosodie par l'énonciateur*

Un aperçu typologique suggère que l'opposition entre analyses autosegmentales-métriques *(tone sequence models)* et modèles d'interaction de facteurs de plusieurs niveaux *(interaction models)* n'est pas uniquement une querelle de théoriciens : l'un et l'autre modèle seraient plus ou moins adéquats à telles ou telles langues. Autrement dit, ces deux modèles

---

[1] Sur un plan très abstrait, ce constat confirme que le fait d'opérer avec des catégories binaires n'est nullement, en lui-même, garantie d'adéquation descriptive ; voir Fant 1971:175, Householder 1971:167, ainsi que les réflexions informelles de Martinet 1993 sur le binarisme.

[2] Cette observation rejoint celle de Di Cristo au sujet du *downstep*, phénomène présent dans certaines langues à tons, et postulé en anglais (et dans de nombreuses autres langues) dans les analyses autosegmentales-métriques de l'intonation : « En réalité, le statut linguistique du downstep ne semble pas clairement établi en dehors des langues à tons » (Di Cristo 2004:95 ; ce point de vue s'oppose à celui exposé dans Ladd 1993, « In defense of a metrical theory of intonational downstep »). Di Cristo évoque la remise en cause par Audra Dainora de la pertinence de la notion de downstep dans la description de l'intonation anglaise (Dainora 2001a, « Eliminating downstep in prosodic labeling of American English ») ; il convient néanmoins de souligner que le détail de l'analyse d'Audra Dainora n'emporte pas nécessairement l'adhésion, du fait que certains paramètres ne paraissent pas avoir été pris en compte (en particulier la ligne de déclinaison, et la fréquence fondamentale intrinsèque et co-intrinsèque). Les tableaux de données fournis (dans l'article cité, et dans la thèse de l'auteur : Dainora 2001b) ne permettent pas au lecteur de se prononcer avec certitude sur la validité des conclusions statistiques.

correspondraient à deux modes d'organisation prosodique ; chaque langue particulière se situerait en un point du continuum entre les deux pôles typologiques correspondants, le pôle des *langues à prosodie calculée*, et le pôle des *langues à prosodie contrôlée par l'énonciateur*. Il semblerait que les langues dont l'organisation prosodique se rapproche le plus clairement du premier type soient essentiellement des langues africaines ; principalement des langues à tons, mais non exclusivement. Dans ces langues, la réalisation phonétique d'un ton est presque entièrement déterminée par la séquence tonale dans laquelle il figure ; cette réalisation ne reflète pas le degré de saillance informationnelle du mot avec lequel le ton se trouve associé. Plus une langue est proche du type opposé, celui des *langues à prosodie contrôlée par l'énonciateur*, plus l'énonciateur est libre de jouer d'un ensemble de morphèmes intonatifs pour prêter à l'expression son tour définitif. Les trois langues étudiées seraient plutôt des *langues à prosodie contrôlée par l'énonciateur*, l'anglais et le vietnamien plus encore que le naxi, cette dernière langue présentant quelques éléments de calcul tonal : tons flottants, et possibilité d'ajout de ton haut à des fins intonatives.

Cette perspective typologique permet de décrire sans paradoxes la variété des systèmes prosodiques, configurations qui résultent de combinaisons variées entre traits prosodiques. Elle ouvre sur de nouveaux projets d'approches expérimentales, dont les résultats ne manqueront pas de la compléter.

Table des matières du Glossaire

Accent 233
Accent contrastif ; accent de focalisation ; accent de phrase 234
Accent d'insistance 234
Accentuation 235
AS 235
Consultant 236
Craquée (voix craquée) 236
Déclinaison 236
DECPA 236
Downstep 237
Electroglottographie 237
F_0 237
F_1, F_2, F_3 237
FK 237
Focalisation contrastive 237
Focus 238
Glottalisation 238
Groupement intonatif 239
H, M, L : abréviation de ton haut, ton moyen, ton bas 239
Insistant (condition I : condition de lecture insistante) 239
Intensification intonative 240
Intonation 240
Intonème 240
L : abréviation de Low (ton bas) 240
Laryngalisation 241
M : abréviation de Moyen (ton moyen). 241
Marquage intonatif du focus 241
Mécanisme phonatoire 241
Mélodie 242
Mise en valeur intonative 243
Morphème intonatif 243
NL 244
O_q : abréviation de quotient ouvert (voir ce terme) 244
Paramètre 244
Prosodie 244
Qualité de voix 245
Quotient ouvert (indication sur la qualité de voix) 245
Registre (langues à registres de qualité de voix) 246
Segmental et suprasegmental 246
Soigné (condition S = condition de lecture soignée) 246
Ton, ton ponctuel, ton modulé 246
Ton ponctuel 247
Ton modulé 247
Tonème 248

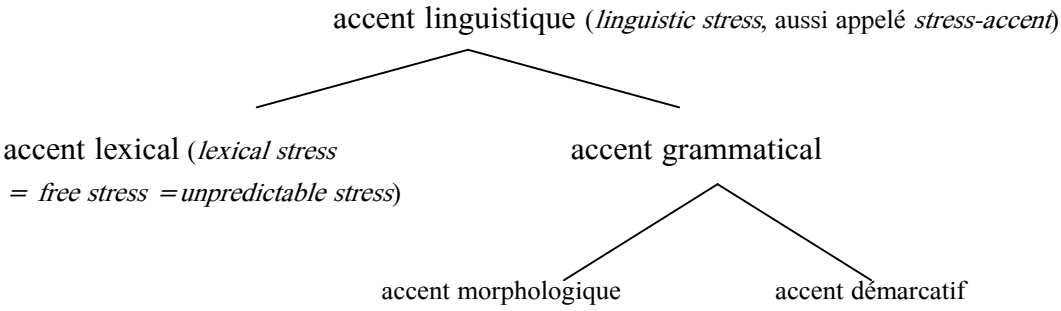
Glossaire et explication des abréviations

Ce bref glossaire vise à faciliter la lecture du présent travail. Parmi les principales notions employées dans le présent travail, beaucoup sont employées dans des sens différents selon les auteurs (ainsi de *prosodie, intonation, accent, ton*) ; il paraissait donc utile de fournir des éclaircissements, sans prétendre offrir un jeu complet de définitions pour la description prosodique, ce qui nécessiterait de distinguer plusieurs niveaux (fonctionnel/phonologique, physiologique, acoustique, perceptif) et de passer en revue les diverses définitions proposées à ce jour ; entreprise qui dépasse le cadre du présent travail.

Lorsque de proches équivalents existent en anglais, ils sont signalés à la fin de l'entrée correspondante.

Accent

A la suite de Larry Hyman (Hyman 1977, « On the nature of linguistic stress »), le terme « accent » est employé pour désigner les réalités suivantes :



> In any given language stress may function in any one or more of these, and in fact, the typical case is that a language with lexical stress may also have some grammatically predictable stress, and that a language with demarcative stress may be occasionally upset by morphological considerations. The final stress in *terminó*, for instance, can be grammatically predicted on the basis of the *–ó* suffix meaning 'third person singular past tense', which always takes stress in Spanish. Other stress oppositions are lexically defined in Spanish (e.g. *término* 'term'), while the general (demarcative) pattern is penultimate stress (e.g. *termíno*, 'I terminate'). (Hyman 1977:39)

Pour plus de détails, voir le chapitre IV, section 3.1.1.

Soulignons que l'**accent d'insistance** (voir ce terme) ne fait pas partie des phénomènes accentuels ainsi définis, mais est de nature intonative : il ressortit à l'intonation pragmatique. L'appellation « accent d'insistance » est néanmoins maintenue ici par commodité, du fait qu'elle est consacrée par l'usage (notamment Coustenoble et Armstrong 1937, Carton, Hirst

et al. 1976) ; néanmoins, à l'avenir, il nous paraîtrait avisé de proposer une désignation de ce phénomène qui évite le terme d'*accent*.

Accent contrastif ; accent de focalisation ; accent de phrase

Dans le souci de réserver le mot d'accent à l'accent lexical et grammatical (voir l'entrée **Accent**), les expressions « accent contrastif », « accent de focalisation » et « accent de phrase » ne sont pas employées dans la présente thèse. Les phénomènes désignés comme « accent contrastif » et « accent de focalisation » par certains auteurs sont ici désignés comme **focalisation contrastive ;** voir ce terme. Nous n'entrerons pas ici dans une description des emplois du terme « accent de phrase » ; signalons simplement l'observation de Nina Grønnum et Jacqueline Vaissière selon laquelle le phénomène en question, bien attesté en anglais, ne serait pas présent en danois et en français, langues qui ne posséderaient pas d'« accent de phrase » (Grønnum 1998b).

Equivalent anglais : *contrastive stress* pour les deux premiers termes, *sentence stress* (parfois *sentence accent*) pour le troisième.

Accent d'insistance

Forme extrême d'**intensification intonative** (voir ce terme) que l'énonciateur attribue à une syllabe ou un mot.

L'**accent d'insistance** est distinct du **marquage intonatif du focus** et de la **focalisation contrastive**[1].

Il importe en outre, en naxi, de distinguer l'**intensification intonative** du phénomène catégoriel d'ajout de ton H (décrit au chapitre II, 2e volet, section 2.4).

En anglais, où il existe deux mots, *stress* et *accent*, il est envisageable d'employer le premier pour l'accent lexical, et le second pour l'**intensification intonative** et l'**accent d'insistance** (« accent intonatif »), à la suite de Dwight Bolinger (Bolinger 1958). Ce n'est pourtant pas ce qui se passe dans l'usage courant, comme en témoigne par exemple l'expression « *contrastive stress* » pour la focalisation contrastive, et « *emphatic stress* » pour l'accent d'insistance, phénomènes intonatifs et non lexicaux ; les définitions proposées par Mary Beckman (Beckman 1986, *Stress and Non-stress Accent*) vont jusqu'à inverser ces deux définitions.

Equivalent anglais : *emphatic stress*, parfois *accent*.

---

[1] Cet usage s'écarte de celui de Mario Rossi, qui définit l'« accent de focalisation » comme « morphème intonatif qui place au premier plan un item de l'énoncé », signalant que ce marqueur de focalisation est « improprement appelé accent d'insistance » (Rossi 1999:10, 203).

Accentuation

L'**accentuation** est entendue ici dans un sens restrictif, qui inclut exclusivement les phénomènes distinctifs lexicaux que sont (selon les langues) l'accent, les accents mélodiques, les tons, ou les registres de qualité de voix.

Inclure dans l'*accentuation* les tons lexicaux (qui existent dans des langues comme le vietnamien) et les registres de qualité de voix (qui existent dans des langues comme le môn) a quelque chose de paradoxal, « accentuation » évoquant immédiatement « accent », et non « ton », encore moins « registre de qualité de voix ». Ce choix résulte de l'extension par analogie de la catégorie « accentuation » du cadre définitoire proposé par M. Rossi et J. Vaissière ; il fait ressortir la caractéristique commune à l'accent lexical (accent libre), aux tons et aux registres de qualité de voix, qui est de posséder un statut lexical distinctif. Il pourrait être indiqué de choisir un autre nom pour la catégorie en question, mais cela n'a pas paru avisé au stade actuel du travail, car cela aurait brouillé les correspondances de vocabulaire avec les cadres théoriques choisis pour référence.

Cette définition du terme d'**accentuation** s'écarte de celle retenue par François Dell (Dell 1984), et par Paul Garde (Garde 1968), chez qui l'**accentuation** est une propriété du morphème, et l'accent une propriété du mot, l'accent d'un mot étant *la réalisation des virtualités accentuelles des morphèmes qui le composent* (proposition mise en perspective par Rossi 1999:31).

Equivalent anglais proposé : la périphrase *lexically distinctive suprasegmentals*, ou (pour éviter **suprasegmental** ; voir ce terme) *lexically distinctive stress, tone or pitch accent*. Ces périphrases paraissent plus éclairantes pour un lecteur de tradition anglo-saxonne que la simple expression « *lexical accentuation* ».

AS

Abréviation du nom du village de /ā ʂə˞/, hameau de /lé pʏ̄ lò/ (en chinois : jin^{1}shan^{1} xiang^{1}, wen^{2}hua^{4} xing^{2}zheng^{4}cun^{1}, leng^{3}bu^{4}luo^{2} zi^{4}ran^{2}cun^{1} 丽江, 金山乡, 文化行政村, 冷不罗自然村), et du dialecte naxi qui y est parlé. Cette localité se situe à environ 20 km au nord-est de la ville de Lijiang (voir cartes 1 et 2, dans l'Annexe 1).

Consultant

Le terme **consultants** désigne les locuteurs qui ont pris part à l'enquête, et fourni les informations sur lesquelles repose le travail. L'emploi de ce terme, à l'origine réservé aux membres de professions prestigieuses (en particulier médecins et juristes) qui donnent des consultations, vise à faire ressortir leur irremplaçable expertise linguistique, leur rôle de premier plan dans la recherche menée sur leur langue. Le terme d'*informateur*, très couramment employé, a des connotations négatives du fait de son emploi, comme synonyme d'*indicateur*, *balance*, pour désigner un membre de la pègre qui fournit des renseignements à la police, traître en cela à son groupe ; le souhait du linguiste est naturellement d'éviter que ses *consultants* soit ainsi perçus par les autres locuteurs de la langue étudiée.

En anglais : *consultant* ou *language consultant*.

Craquée (voix craquée)

Voir **Mécanisme phonatoire**.

En anglais : *creaky voice*.

Déclinaison

Décroissance progressive de F_0 au cours de l'énoncé affirmatif, l'écart se réduisant entre ligne de base et ligne de crête (voir la section 4.1 de l'Introduction).

Par extension, le terme est employé pour désigner la décroissance d'autres paramètres : intensité acoustique, ampleur des mouvements des articulateurs (voir par exemple Vayra et Fowler 1992, « Declination of supralaryngeal gestures in spoken Italian »).

La déclinaison comporte une composante physiologique, dont l'importance demeure disputée, et une composante linguistique, le *jeu sur la ligne de déclinaison* (son amplification, son annulation, sa remise à zéro) étant porteur d'informations linguistiques. (A ce sujet, voir notamment les réflexions d'Ilse Lehiste : Lehiste 1975, 1979, 1982.)

En anglais : *declination*.

DECPA

Derivative-Electroglottographic Closure Peak Amplitude : amplitude des pics positifs sur la dérivée du signal électroglottographique, correspondant au pic de vélocité de l'augmentation de surface d'accolement des plis vocaux lors de la fermeture glottique. Le terme est dû à Nathalie Henrich (communication personnelle). Au sujet de ce paramètre, voir l'Annexe 2, section 3 (ainsi que Michaud 2004b, « A Measurement from Electroglottography: DECPA, and its Application in Prosody »).

Downstep

La notion de *downstep* est définie de façon différente selon les auteurs, d'où le risque de confusions. Le mot est employé ici dans le sens que retient Annie Rialland au terme d'un aperçu historique sur cette notion (Rialland 1997, « Le parcours du "downstep", ou l'évolution d'une notion » ; voir également Connell 2001 et références citées). Le *downstep automatique* (parfois appelé *downdrift*, dénomination non retenue ici) est le phénomène par lequel, dans les séquences HLHLHL, les H se succèdent à une hauteur chaque fois diminuée. Le *downstep non automatique* est un phénomène d'abaissement de tons H successifs, déclenché par des tons L flottants (voir Rialland 2001:303 et références citées).

Electroglottographie

Méthode expérimentale (non invasive) de mesure de la surface d'accolement des plis vocaux, inventée par Fabre (1957). Un courant de haute fréquence et de très faible intensité circule entre deux électrodes placées de part et d'autre du larynx. Voir chapitre I, section 1.

F_0

F_0, ou fréquence fondamentale : fréquence de vibration des plis vocaux au cours de la phonation. F_0 est le premier harmonique du signal (H_1).

F_1, F_2, F_3

Premier, deuxième et troisième formant, respectivement[1]. Les formants, régions d'harmoniques renforcés, véhiculent des indications sur la configuration des cavités supraglottiques (Fant 1960), et par là (de façon indirecte) sur la position des articulateurs. Cette numération pose problème dans certains cas (notamment dans le cas des voyelles nasales) ; voir chapitre I, section 2.1.

FK

Abréviation du nom du village de /fɣ̩ kʰō/ (en chinois : feng^{4}ke^1 xiang^1, shan^{4}mei^3 xing^{2}zheng^{4}cun^1, heng^{2}ke^3 er^4 zu^3 丽江, 奉科乡, 善美行政村, 恒可二组), et du dialecte naxi qui y est parlé. Ce village se situe à la frontière nord-nord-est de la préfecture de Lijiang (voir carte 1, dans l'Annexe 1).

Focalisation contrastive

La **focalisation contrastive** est couramment représentée à l'écrit en mettant en majuscules le mot contrasté : pour prendre des exemples caricaturaux mais parlants, en anglais, *MARY ate*

[1] F est également employé comme abréviation de Féminin : pour faciliter le classement des données, chacun des consultants ayant participé aux enquêtes est associé à une lettre (F pour Féminin, M pour Masculin) et un numéro, dans l'ordre où ils ont participé à l'enquête : par exemple, F1 pour la première locutrice.

it et *Mary ATE it* seront, l'un une réponse à *Who ate my chocolate bar?*, l'autre une réponse à *What did Mary do with my chocolate bar?* Ce phénomène paraît correspondre à ce que Rossi 1999:203, dans sa description du français, définit comme l'*accent de focalisation*, « morphème intonatif qui place au premier plan un item de l'énoncé ».

La distinction entre **morphème intonatif de focalisation** (*marquage intonatif du focus)* et **focalisation contrastive** n'est pas aisée à établir en pratique dans tous les cas ; néanmoins, c'est un fait typologique important que la focalisation contrastive, qui existe dans des langues comme l'anglais, le chinois (voir en particulier Marie-Claude Paris [Paris 1999:206-207]) ou le vietnamien, est absent de beaucoup d'autres langues, dont le naxi, qui possèdent néanmoins un marquage intonatif du focus.

Dans de nombreuses études, la **focalisation contrastive** est appelée *accent contrastif* ou *accent de focalisation*, termes qui attirent l'attention sur la réalisation phonétique du phénomène, laquelle présente certaines ressemblances avec la réalisation de certains types d'accent (voir ce mot). Dans le souci de réserver le mot d'accent à l'accent lexical et grammatical (voir l'entrée **Accent**), les expressions « accent contrastif » et « accent de focalisation » ne sont pas employées dans la présente thèse.

Equivalent anglais : *contrastive stress*.

Focus

Le terme de **focus** s'entend ici dans le cadre d'un jeu de question-réponse : le focus est *ce sur quoi porte la question* (dans une question), *ce qui apporte une réponse à la question posée* (dans une réponse). Cette caractérisation laisse ouvertes diverses questions, mais suffit pour le propos de la présente étude. (Pour un aperçu plus large, nous renvoyons à Knud Lambrecht [Lambrecht 1994], Gussenhoven 1983, et un article à paraître de Gussenhoven : « Types of Focus in English » ; au sujet du focus dans une langue à tons d'Asie, le chinois mandarin, voir Marie-Claude Paris [Paris 1999:204 et 214] et Xu Yi [Xu Yi 1997, 1999].)

Glottalisation

Le terme « glottalisation » connaît des emplois variables selon les auteurs. Il est ici employé (suivant en cela Henton, Ladefoged *et al.* 1992:73) comme hyperonyme *(cover term)* pour deux phénomènes : la constriction glottale (telle qu'elle se manifeste exemplairement en fin de ton B2 dans les données vietnamiennes), et la laryngalisation, ou passage en voix craquée (telle qu'elle se manifeste très fréquemment en fin de ton C1 dans les données vietnamiennes) ; voir chapitre III, section 3.3.3.1 et figures 3.57 et 3.58.

En anglais : *glottalisation*.

Groupement intonatif

Il n'a pas paru nécessaire, pour le propos de la présente étude, d'aborder de front la question des distinctions à établir entre diverses catégories de groupements intermédiaires entre le mot et l'énoncé, tels que « mot intonatif », « syntagme intonatif », « groupe intonatif ». Les termes « groupement intonatif » et « unité intonative » sont ici employés (de façon interchangeable) dans un sens large, comme **hyperonyme pour les divers types de groupements intermédiaires entre le mot et l'énoncé**.

Une hiérarchie des unités intonatives est avancée par Elisabeth Selkirk, qui propose de distinguer, de l'unité la plus large à la plus petite : *intonational phrase, phonological phrase, prosodic word, foot, syllable* (Selkirk 1984, 1986 ; voir également Nespor et Vogel 1986 ; diverses unités intermédiaires—telles que l'*intermediate phrase* de Pierrehumbert et Beckman 1988—ont été proposées ; voir également Hirst et Di Cristo 1998).

Sur la base d'observations sur le français (et plus secondairement sur l'anglais et d'autres langues), Jacqueline Vaissière propose d'employer les unités suivantes : le paragraphe oral, unité supérieure à l'énoncé ; l'énoncé ; le groupe de souffle *(breath-group)*[1] *;* le groupe mélodique *(melodic phrase)*, qui se caractérise par un retour final à la ligne de base ; le groupe prosodique *(prosodic phrase)* ; et le mot prosodique *(prosodic word).* (Voir notamment Vaissière 1980, et Vaissière et Michaud à paraître.)[2]

H, M, L : abréviation de ton haut, ton moyen, ton bas

Les tons ponctuels sont désignés comme H, M et L (abrégeant *High, Mid, Low*), non H, M, B *(Haut, Moyen, Bas)* pour éviter la confusion avec la catégorie tonale B du système de notation A, B, C, D en usage dans la description des langues d'Asie du Sud-Est (système qui à notre connaissance est dû à Gedney 1986, 1989).

Insistant (condition I : condition de lecture insistante)

Les deux conditions de lecture employées dans l'expérience rapportée au chapitre III sont décrites en détail dans les sections 1.1.2, 2.2 et 3.2 ; les appellations « condition de lecture soignée » (abrégée en S dans les tableaux et figures) et « condition de lecture insistante » (abrégée en I dans les tableaux et figures), choisies pour leur concision, n'ont pas valeur définitoire.

---

[1] De même que la **déclinaison** a une base physiologique mais connaît des emplois linguistiques (jeu linguistique sur la déclinaison), le **groupe de souffle** a une base physiologique mais ne coïncide plus systématiquement, en parole, avec une prise de souffle.

[2] Signalons également, dans une autre perspective, une thèse récente qui part de données de parole « spontanée » pour construire une réflexion sur la hiérarchisation des constituants discursifs : Passot 2004.

Intensification intonative

Phénomène scalaire de mise en valeur d'une syllabe par des moyens intonatifs, possible à divers degrés dans toutes les langues, et dont l'**accent d'insistance** représente une forme extrême. L'intensification intonative est une composante de l'intonation pragmatique.

(Certains auteurs parlent de « soulignement intonatif » [par exemple A. Rialland, M.-A. Morel et L. Danon-Boileau].)

Equivalent anglais proposé : *intonational intensification*, ou plus simplement *emphasis*.

Intonation

L'intonation est une des composantes de la prosodie, cette dernière incluant l'accentuation, l'intonation, et les facteurs de performance dont le rythme. (Précisons que ce jeu de définitions, qui se fonde sur les travaux de M. Rossi et J. Vaissière, n'est pas unanimement reçu. Une formulation très proche—nous semble-t-il—est proposée par Albert Di Cristo : voir Di Cristo 2004:183, Di Cristo 1998.)

L'intonation se divise en une composante morpho-syntaxique (découpage en constituants intonatifs) et une composante pragmatique, désignées par commodité comme *intonation syntaxique* et *intonation pragmatique* (voir l'Introduction, section 4) ; elle comprend également l'expression des attitudes et émotions. « Intonation syntaxique » et « intonation pragmatique » correspondent respectivement à la fonction démarcative et la fonction d'emphase d'Ivan Fónagy (Fónagy 2003, « Des fonctions de l'intonation : essai de synthèse » ; voir également Fónagy 1983, 1989).

Equivalent anglais : le terme *intonation* prête à confusion du fait que certains auteurs l'emploient dans le sens que possède ici *prosodie*. Cet emploi est dû, en dernière analyse, à l'absence de distinction entre le plan linguistique (ou, dit autrement, *fonctionnel*, ou encore *phonologique*) et celui des réalisations phonétiques, la « courbe intonative » d'un énoncé étant simplement sa courbe de hauteur : ainsi chez Yetunde Laniran, « The phonetic study of intonation (that is, fundamental frequency realization) (..) [Laniran 1992:2] ».

Intonème

Voir **morphème intonatif**.

L : abréviation de Low *(ton bas)*

Les tons ponctuels sont désignés comme H, M et L (abrégeant *High, Mid, Low*), non H, M, B *(Haut, Moyen, Bas)* pour éviter la confusion avec la catégorie tonale B du système de notation A, B, C, D en usage dans la description des langues d'Asie du Sud-Est (système qui à notre connaissance est dû à Gedney 1986, 1989).

Laryngalisation

Passage en voix **craquée** (désignée ici comme **mécanisme phonatoire 0** ; voir **Mécanisme phonatoire**).

Equivalent anglais proposé : *laryngealisation = lapse into creaky voice*.

M : abréviation de Moyen (ton moyen)[1].

Marquage intonatif du focus

Dans les langues qui possèdent un **marquage intonatif du focus**, le statut informationnel de **focus** (voir ce terme) se reflète dans des modifications intonatives de la réalisation du mot concerné.

Anglais, vietnamien et naxi présentent un marquage intonatif du focus. L'existence de langues sans marquage intonatif du focus est établie par les études sur le wolof (Rialland et Robert 2001, 2003).

Equivalent anglais proposé : *intonational marking of focal status*.

Mécanisme phonatoire

En voix parlée comme en voix chantée, un locuteur peut produire le voisement en *voix de tête* (dans le haut de sa tessiture), en *voix de poitrine* (dans le bas de sa tessiture), ou en *voix craquée* (généralement tout en bas de sa tessiture), mode de vibration des plis vocaux qui est irrégulier et de très basse fréquence. Ces phénomènes sont depuis longtemps l'objet d'études (Catford 1964, « Phonation types: The classification of some laryngeal components of speech production » ; Catford 1977:95-106 ; Laver 1980, *The phonetic description of voice quality ;* Titze 1994, 1995 ; Ladefoged et Maddieson 1996:73). La voix craquée est parfois appelée *friture vocale* (Hollien, Moore *et al.* 1966, « On the nature of vocal fry ») ou *laryngalisation*.

Dans la description des trois mécanismes phonatoires (ou *mécanismes laryngés*) que sont la « voix craquée », la « voix de poitrine » et la « voix de tête », il a paru utile de mentionner les termes proposés par Bernard Roubeau sur la base d'études électroglottographiques de la voix chantée (Roubeau, Chevrie-Muller *et al.* 1987, 1991) :

- le mécanisme 0 ou voix craquée *(creaky voice)*[2]. Le passage en mécanisme 0 est désigné comme **laryngalisation**.

---

[1] M est également employé comme abréviation de Masculin : pour faciliter le classement des données, chacun des consultants ayant participé aux enquêtes est associé à une lettre (F pour Féminin, M pour Masculin) et un numéro, dans l'ordre où ils ont participé à l'enquête : par exemple, M1 pour le premier locuteur.

[2] Certains auteurs emploient les termes de *vocal fry* et *vocal pulse* pour désigner un type de laryngalisation dans lequel une fréquence demeure discernable, opposant ce type de vibration au *creak (creaky voice)*, dans

- le mécanisme 1 correspond grossièrement à ce qui est communément nommé « voix de poitrine »
- le mécanisme 2 coïncide pour l'essentiel avec la « voix de tête ».

Signalons que B. Roubeau décrit également un mécanisme 3 ou « voix de sifflet », employé dans la partie la plus élevée (suraiguë) de la tessiture. Le mécanisme 3 n'est pas abordé ici, n'étant pas représenté dans notre corpus.

A ce jour, l'usage de ce mode de classification ne s'est pas imposé dans la communauté des chercheurs ; il est néanmoins employé ici (tout en rappelant, dans le corps du texte, les termes plus traditionnels que le cadre de B. Roubeau se propose d'affiner), car il nous paraît aussi éclairant pour la voix parlée que pour la voix chantée, et peut assez commodément être mis en rapport avec les données électroglottographiques : voir un exemple au chapitre I, section 1.2.3. Une notation par des numéros est en elle-même moins parlante que les expressions telles que « voix de tête », mais elle est en fait plus précise, dans la mesure où les termes « voix de tête » et « voix de poitrine » s'inscrivent, dans l'emploi qui en est traditionnellement fait dans le domaine de la voix chantée, à l'intérieur d'un paradigme qui comporte également d'autres catégories, telles que la « voix mixte », qui ne correspondent en fait pas à un mécanisme laryngien distinct, mais sont le fruit d'un entraînement des chanteurs (voir Castellengo, Chuberre *et al.* 2004, « Is Voix Mixte, the vocal technique used to smoothe the transition across the two main laryngeal mechanisms, an independent mechanism? »

La voix soufflée n'est pas considérée ici comme un mécanisme phonatoire distinct : il s'agit, selon les cas, d'une phonation en voix craquée (mécanisme 0), « voix de poitrine » (mécanisme 1) ou « voix de tête » (mécanisme 2) ; le fait que le débit d'air soit élevé ne constitue pas un changement de mécanisme phonatoire, seulement de **qualité de voix**.

Un équivalent anglais existe : *laryngeal mechanism*.

Mélodie

Perception qu'a l'auditeur du paramètre de fréquence fondamentale (définition rappelée par Rossi 1999:7, 207).

L'équivalent anglais *pitch* est souvent employé indistinctement pour la fréquence fondamentale et la mélodie. L'expression *pitch contour* est très couramment employée pour désigner la courbe de fréquence fondamentale : ainsi, dans le manuel du logiciel PRAAT : « To see the pitch contour of an existing sound as a function of time, select a Sound or LongSound object and choose Edit ». Cet usage paraît à éviter car le *pitch* est une notion

---

lequel la vibration est « chaotique » (Nathalie Henrich, communication personnelle). Les données que nous avons observées ne paraissent pas nécessiter une telle distinction ; ces termes sont ici employés de façon interchangeable.

perceptive, non acoustique[1]. De même en français : un logiciel peut extraire une courbe de fréquence fondamentale, non une *courbe mélodique*.

Mise en valeur intonative

Voir *intensification intonative*.

Morphème intonatif

Afin de ne pas confondre le plan des catégories phonologiques et celui des réalisations phonétiques (autrement dit, le plan de la forme linguistique et celui de la substance phonique), Rossi distingue **morphème intonatif** et **intonème**. Les primitives de l'intonation sont des **morphèmes intonatifs :** morphèmes d'intonation syntaxique, morphèmes d'intonation sémantique-pragmatique, et morphèmes porteurs d'expressivité (les *clichés mélodiques* de Fónagy). Ces morphèmes se réalisent par des **intonèmes**, « traits construits sur l'espace à 3 dimensions (fréquence fondamentale, intensité, temps) » (Rossi 1999:9, 207).

Sur le fond, cette distinction est reprise dans le présent travail. S'agissant des termes employés, **intonème** nous paraît évoquer le plan des catégories phonologiques (par analogie avec le **phonème**, unité distinctive abstraite, à la différence du son, parfois appelé **phone**, par lequel il se réalise ; ou avec le **tonème** (voir ce terme), qui permet pareillement de désigner une unité distinctive sans préjuger de sa réalisation phonétique. Cette analogie conduirait à faire du terme **intonème** un synonyme de **morphème intonatif**[2].

---

[1] Une discussion a récemment eu lieu à ce sujet sur le forum du logiciel PRAAT. Paul Boersma maintient néanmoins l'emploi du mot *pitch*, pour les raisons suivantes : « ... some of the pitch algorithms yield an acoustic fundamental frequency and some yield a more perceptually-based result. Even the "acoustic" pitch algorithms in PRAAT have many features that make them work like human pitch perception rather than like mathematical periodicity: the robustness against additive noise, the bias toward the higher pitch candidate if every even cycle is a bit stronger than every odd cycle, the capability of smoothing a pitch path through locally second-best pitch candidates, and not hearing any periodicity in silent stretches. In that sense most of PRAAT's pitch algorithms are perceptually-oriented (…). F_0 and pitch are *quasi*. » Suite à ce message, diffusé à tous les abonnés de la liste de diffusion, Plinio Almeida Barbosa (Université de Campinas, Brésil) a répondu à Paul Boersma et à nous dans les termes suivants : « I really DO NOT want to cause additional misunderstanding (that is why I am not copying for the whole PRAAT-Group list), but I totally subscribe to what you are saying, Alexis. I think everyone who teaches speech science (I think it is the case of you too, Paul) repeats these very facts to one's students. I do think F_0 should be substituted for pitch too, since the latter cannot be grasped without a perception test using humans ('cause it is a sensation). But I understanding there is a 'tradition' of using the term in speech synthesis/recognition circles, for instance. »

[2] Tel est l'emploi que paraît en faire A. Di Cristo (séminaire de doctorat, 2002) lorsqu'il divise la « prosodologie » en « accentologie, tonologie, intonologie et chronématologie », les *intonèmes* se plaçant au même niveau que les *accentèmes*, *tonèmes* et *chronèmes*.

Plutôt que d'employer le terme d'**intonème**, nous parlons donc de **morphème intonatif** et de **réalisation (phonétique)** du morphème intonatif.

NL

Abréviation du nom du village de /ndū lè/ (en chinois : jin[1]shan[1] xiang[1], gui[4]feng[1] da[4]lai[2] xing[2]zheng[4] xia[4] cun[1] 丽江, 金山乡, 贵峰大来行政下村 /大来二村), et du dialecte naxi qui y est parlé. Cette localité se situe à une douzaine de kilomètres au sud de la ville de Lijiang (voir cartes 1 et 2, dans l'Annexe 1).

O_q *: abréviation de* quotient ouvert *(voir ce terme)*

Paramètre

Paramètre est entendu ici au plan technique, celui de la mesure : des *événements* ont lieu au plan de la production ; l'enquêteur collecte des *données* (telles que : données audio, données électroglottographiques, données de débit d'air...) lors de ces événements ; à partir des données, l'enquêteur calcule certains *paramètres* tels que fréquence fondamentale (d'où peuvent être dérivés des paramètres secondaires : pente de F_0, position et valeur d'un maximum local de F_0, durée d'un mouvement de F_0), longueur, intensité, fréquence des formants, quotient ouvert ; liste qui demeure ouverte. Au plan de l'analyse statistique, on parlera de *facteurs* et non de *paramètres* : par exemple, chez tel locuteur, *le* ***facteur*** *CONDITION DE LECTURE a un effet significatif sur le* ***paramètre*** *qu'est le quotient ouvert.*

Prosodie

Le domaine de la prosodie est ici délimité de façon large, dans l'esprit (nous semble-t-il) des travaux de M. Rossi et J. Vaissière : la prosodie inclut l'*accentuation*, l'*intonation*, et les *facteurs de performance* (dont le rythme et le débit). **L'accentuation est l'ensemble des phénomènes distinctifs lexicaux** que sont (selon les langues) l'accent lexical, les accents mélodiques, les tons ou les registres de qualité de voix ; **l'intonation, domaine d'une grande complexité, inclut les faits de structuration post-lexicale que sont le découpage en constituants (composante *syntaxique* de l'intonation, à fonction *démarcative*) et le marquage de la structure informationnelle (composante *pragmatique* de l'intonation)**, ainsi que l'expression des attitudes et émotions.

Au plan phonétique, la prosodie se manifeste par *toutes les variations sous-glottiques, glottiques et supra-glottiques non prévisibles à partir de la connaissance des phonèmes.* (Précisons que ce jeu de définitions, qui se fonde sur les travaux de M. Rossi et J. Vaissière, n'est pas unanimement reçu. Une formulation très proche—nous semble-t-il—est proposée par Albert Di Cristo : voir Di Cristo 2004:183, Di Cristo 1998.)

Cette définition programmatique ouvre des perspectives de recherche larges, sans clore l'inventaire des faits prosodiques (au plan de la substance non plus qu'au plan de la forme).

Elle s'écarte de l'usage courant dans les publications anglo-saxonnes : ainsi, David Crystal caractérise « prosody » comme « A term used to refer collectively to variations in PITCH, LOUDNESS, TEMPO and RHYTHM », et « intonation » comme « A term referring to the distinctive use of patterns of PITCH, or melody » (Crystal 1985) ; définition qui se place au plan de la substance, et n'établit pas de séparation entre faits d'accentuation et faits intonatifs.

Qualité de voix

Le terme est entendu ici dans le sens de *mode de vibration des plis vocaux*. Cette définition est relativement restrictive : les travaux sur les langues qui opposent plusieurs qualités de voix (aussi appelées registres phonatoires, ou modes phonatoires) montrent que les différences ne sont pas uniquement laryngales : le sphincter aryépiglottique et le pharynx entrent souvent en jeu dans la réalisation de la qualité de voix *tendue* ; le timbre vocalique est également affecté (voir notamment Edmondson, Esling *et al.* 2001, « The aryepiglottic folds and voice quality in the Yi and Bai languages: Laryngoscopic case studies », Esling et Edmondson 2002, « The laryngeal sphincter as an articulator: Tenseness, tongue root and phonation in Yi and Bai », et les réflexions générales de Catford 1964 et Laver 1980).

La place centrale que nous accordons au mode de vibration des plis vocaux dans la présente définition tient en partie au rôle que joue dans notre étude la technique exploratoire de l'électroglottographie, laquelle renseigne sur la surface d'accolement des plis vocaux, non sur les ajustements du sphincter aryépiglottique et du pharynx. Dans la présente thèse, la qualité de voix est en effet étudiée par le biais du **quotient ouvert calculé par électroglottographie**.

Cet emploi de l'expression *qualité de voix* est différent de celui qu'en fait Laver 1980:1, qui l'emploie dans le sens de : « the characteristic auditory coloring of an individual speaker's voice » ; de même chez Story et Titze 2002 (« A preliminary study of voice quality transformation based on modifications to the neutral vocal tract area function »).

Dans son enseignement, Jacqueline Vaissière oppose la **qualité de la voix**, qui recouvre les dimensions sous-glottique et glottique, à la **qualité de la parole**, qui recouvre les dimensions supra-glottiques : contrôle des mouvements des articulateurs (voile du palais, langue, lèvres...).

Equivalent anglais : *voice quality*.

*Quotient ouvert (indication sur la **qualité de voix**)*

Le **quotient ouvert** (O_q) est égal à la durée de la phase ouverte d'un cycle glottique divisée par la durée totale du cycle. Au plan expérimental, la durée de la phase ouverte est ici estimé par la distance qui sépare un pic d'ouverture et le pic de fermeture suivant sur la dérivée du signal électroglottographique ; la durée totale du cycle est estimée par la durée séparant deux

pics de fermeture successifs (voir chapitre I, section 1.2.3, et la figure 1.6, reprise ci-dessous).

Registre (langues à registres de qualité de voix)

Registre est employé ici dans un sens très restrictif, comme forme abrégée de *registre de qualité de voix*. Dans les langues à registres de qualité de voix (appelées ici simplement *langues à registres*), la **qualité de voix** (voir ce terme) a un rôle distinctif au plan lexical. Ainsi, le môn oppose un registre de voix soufflée à un registre de voix modale, ce qui était encore le cas du khmer à date récente (Henderson 1952). Plus encore que d'autres traits linguistiques, les registres de qualité de voix tendent à avoir de multiples corrélats (mode de vibration des plis vocaux, mais aussi longueur plus grande des rimes porteuses d'une qualité de voix non modale, différences dans l'articulation vocalique, différences de fréquence fondamentale) et à connaître une forte variabilité entre locuteurs.

Dans les études prosodiques, le terme de « registre » est fréquemment employé comme synonyme de « niveau de hauteur mélodique », cela soit de façon informelle (remarquant qu'une syllabe est réalisée *dans un registre élevé*, ou *dans un registre grave*, sans que cela implique une division catégorielle), soit pour désigner des niveaux discrets, en particulier dans la description des langues à tons ponctuels (ainsi, les trois principaux tons du naxi pourraient être décrits comme : registre haut, registre moyen, registre bas).

Segmental et suprasegmental

L'opposition très courante entre « segmental » et « suprasegmental » (notamment employée par Lehiste 1970) n'est pas employée ici, notamment du fait que l'articulation des « segments » a part à la prosodie telle qu'elle est définie ici, et que la frontière entre segmental et suprasegmental paraît difficile à tracer. (Voir Rossi 1999:19-32, « La dichotomie *segmental/suprasegmental* est-elle justifiée ? »)

Soigné (condition S = condition de lecture soignée)

Les deux conditions de lecture employées dans l'expérience rapportée au chapitre III sont décrites en détail dans les sections 1.1.2, 2.2 et 3.2 ; les appellations « condition de lecture soignée » (abrégée en S dans les tableaux et figures) et « condition de lecture insistante » (abrégée en I dans les tableaux et figures), choisies pour leur concision, n'ont pas valeur définitoire.

Ton, ton ponctuel, ton modulé

Sauf mention contraire, le terme de **ton** est réservé aux tons lexicalement distinctifs de langues telles que le naxi et le vietnamien *(langues à tons)*. Selon la définition de William Welmers reprise et adaptée par Larry Hyman : « A language with tone is one in which an

indication of pitch enters into the lexical realization of at least some morphemes » (Welmers 1959, 1973 ; Hyman 2001c:1367).

Certains tons sont **ponctuels**, d'autres **modulés**.

Ton ponctuel

Au sens propre, un ton ponctuel (en anglais : *level tone*) est un **tonème** qui est défini uniquement par une caractéristique de hauteur[1]. (Ainsi, le naxi possède trois **tons ponctuels** : H, M et L.) Leur réalisation phonétique (plan de la *substance*) peut comporter des traits secondaires autres que F_0 ; mais leur spécification phonologique se limite à un niveau relatif de hauteur.

Plusieurs **tons ponctuels** associés à une même syllabe se réalisent typiquement comme un contour tonal : ainsi, en naxi, l'ajout d'un ton H à une syllabe au ton L crée un contour montant, /LH/.

Les tons ponctuels sont parfois désignés comme tons-registres *(register tones)*, ou registres tonals *(tone registers)*, car ils se définissent par un registre de fréquence fondamentale : par exemple, en naxi, registre haut, moyen ou bas. Ce terme est évité ici du fait que *registre* est employé comme abréviation de **registre de qualité de voix** (voir ce terme). Par opposition aux tons-contours, les tons ponctuels sont parfois appelés tons plains *(level tones)* ; cette dernière désignation nous paraît avoir le léger inconvénient, en français, de suggérer que la réalisation phonétique du ton concerné serait plane, alors qu'un ton ponctuel bas est généralement réalisé *descendant* au plan phonétique.

Equivalent anglais : *level tone*.

Ton modulé

L'expression de **ton modulé** (qui peut également être appelé ton-contour, veillant à ne pas le confondre avec le *contour tonal* qui résulte de la combinaison de plusieurs **tons ponctuels**) désigne ici un **tonème** qui n'est pas formé d'une séquence de **tons ponctuels**. Kenneth Pike décrit ainsi la distinction entre systèmes de tons modulés (non analysables) et systèmes de tons ponctuels : « the glides of a contour system must be treated as unitary tonemes and cannot be broken down into end points which constitute lexically significant contrastive pitches » (Pike 1948:10). (Par exemple, les tons du tamang (Mazaudon 1973) entrent dans la catégorie des **tons modulés**.) Pike souligne la présence de types intermédiaires, tout en maintenant la nécessité de distinguer les deux catégories.

---

[1] Cet emploi de l'expression *ton ponctuel* ne doit pas être confondu avec l'usage de R.K. Sprigg, qui, dans la représentation stylisée des tons, emploie un simple point pour noter la mélodie de syllabes qu'il juge trop courtes pour porter un contour tonal. Merci à Martine Mazaudon pour nous avoir signalé cette divergence dans l'emploi des termes.

Equivalent anglais : *contour tone ;* il est en fait nécessaire d'employer une expression telle que *unitary contour tone*, qui indique explicitement que ce contour n'est pas décomposable, car le terme « contour » est fréquemment employé pour désigner une combinaison de tons ponctuels.

Tonème

Expression équivalente de *ton lexical*, notamment employée par Kenneth Pike pour souligner que *tonème* et *phonème* partagent la caractéristique de posséder un *statut distinctif au plan lexical*, ainsi que pour distinguer l'entité abstraite qu'est le **tonème** de ses corrélats au plan de la substance (c'est-à-dire au plan de sa réalisation phonétique), qui peuvent être variés, incluant des variations dans l'articulation consonantique et vocalique, des différences d'intensité, de qualité de voix, de longueur, et non pas seulement de fréquence fondamentale comme pourrait laisser entendre le terme **ton**, identifié avec la seule caractéristique de *hauteur* dans de nombreux travaux linguistiques.

Dans le présent travail, par souci de simplicité, le terme de **tonème** n'est pas employé : notre choix consiste à employer le terme **ton** pour désigner le phénomène en question.

Equivalent anglais : *toneme*.

Références

Cette liste couvre les ouvrages cités dans les Annexes, aussi bien que les ouvrages cités dans le corps de la thèse.
Les références en chinois sont fournies en fin de bibliographie.

Abbi A., 1997, Le redoublement dans les langues d'Asie du sud, *Faits de langues*, 10, pp. 31-36.

Abramson A. S., 1978, Static and dynamic acoustic cues in distinctive tone, *Language and Speech*, 21, pp. 319-325.

Abramson A. S., 1979, The noncategorical perception of tone categories in Thai, in *Frontiers of speech communication research*, B. Lindblom et S. Öhman, London, Academic Press, pp. 127-134.

Abramson A. S., 2004, The plausibility of phonetic explanations of tonogenesis, in *From traditional phonology to modern speech processing: Festschrift for Professor Wu Zongji's 95th birthday*, G. Fant, H. Fujisaki, Cao Jianfen et Xu Yi, Beijing, Foreign Language Teaching and Research Press, pp. 17-29.

Abramson A. S. et Lisker L., 1970, Discriminability along the voicing continuum: Cross-language tests, 6th International Congress of Phonetic Sciences, Prague, pp. 569-573.

Abramson A. S., Thongkum T. L. et Nye P. W., 2004, Voice register in Suai (Kuai): An analysis of perceptual and acoustic data, *Phonetica*, 61(2-3), pp. 147-171.

Adamczewski H. et Delmas C., 1982, *Grammaire linguistique de l'anglais*, Collection U, Paris, Armand Colin.

Adank P., van Hout R. et Smits R., 2004, An acoustic description of the vowels of Northern and Southern Standard Dutch, *Journal of the Acoustical Society of America*, 116(3), pp. 1729-1738.

Akinlabi A., 1985, *Tonal Underspecification and Yorùbá Tone*, Ph. D., University of Ibadan.

Allen G., 1978, Vowel duration measurement: a reliability study, *Journal of the Acoustical Society of America*, 63(4), pp. 1176-1185.

Allen J. S. et Miller J. L., 1999, Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words, *Journal of the Acoustical Society of America*, 106, pp. 2031-2039.

Amelot A., 2004, *Etude aérodynamique, fibroscopique, acoustique et perceptive des voyelles nasales du français*, thèse de doctorat, Laboratoire de Phonétique et Phonologie, Paris, Université Paris 3.

Amelot A., Roubeau B., Crevier-Buchman L. et Maeda S., 2004, Prise de données simultanées aérodynamiques et fibroscopiques durant la production des voyelles nasales : comparaison avec des données prises séparément, XXVe Journées d'Etude de la Parole, Fès (Maroc).

Anderson S. R., 1985, *Phonology in the Twentieth Century: Theories of Rules and Theories of Representation*, Chicago, London, University of Chicago Press.

Andruski J. E. et Costello J., 2004, Using polynomial equations to model pitch contour shape in lexical tones: an example from Green Mong, *Journal of the International Phonetic Association*, 34(2), pp. 125-140.

Ao B., 1990, Lip rounding and formant frequencies of the Nantong vowel system, *Journal of the Acoustical Society of America*, 87(S1), p. S122.

Arvaniti A., Ladd R. et Mennen I., 1998, Stability of tonal alignment: the case of Greek prenuclear accents, *Journal of Phonetics*, 26, pp. 3-25.

Assadi S., 2003, Laryngographic, acoustic and perceptual study of glottalization phenomena in Persian, International Congress of Phonetic Sciences, Barcelona, pp. 3117-3120.

Astesano C., 1999, *Rythme et discours: invariance et sources de variabilité des phénomènes accentuels du français*, thèse de doctorat, Université de Provence.

Atal B. S., Chang J. J., Mathews M. V. et Tukey J. W., 1978, Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique, *Journal of the Acoustical Society of America*, 63(5), pp. 1535-53.

Auran C., 2004, *Prosodie et anaphore dans le discours en anglais et en français: cohésion et attribution référentielle*, thèse de doctorat, Aix-Marseille, Université de Provence.

Auran C., Bouzon C. et Hirst D., 2004, The Aix-MARSEC Project: An Evolutive Database of Spoken British English, Speech Prosody 2004, Nara, Japan, pp. 561-564.

Bacot J., 1913, *Les Mosso, ethnographie des Mosso, leurs religions, leur langue et leur écriture, avec les documents historiques et géographiques relatifs à Li-kiang par Ed. Chavannes*, Leiden, E.J. Brill.

Baken R. J., 1992, Electroglottography, *Journal of Voice*, 6(2), pp. 98-110.

Basbøll H., 1999, Syllables in Danish, in *The syllable: views and facts*, H. van der Hulst et N. Ritter, Berlin, Mouton de Gruyter, pp. 69-92.

Basset P., Amelot A., Vaissière J. et Roubeau B., 2001, Nasal airflow in French spontaneous speech, *Journal of the International Phonetic Association*, 31(1), pp. 87-100.

Baxter W. H., 1992, *A handbook of Old Chinese phonology*, Trends in Linguistics Studies and Monographs 64, Berlin, Mouton de Gruyter.

Beckman M., 1986, *Stress and Non-stress Accent*, Netherlands Phonetic Archives No. 7, Dordrecht, Foris.

Beckman M. et Edwards J., 1990, Lengthening and shortening and the nature of prosodic constituency, in *Laboratory Phonology I*, J. Kingston et M. Beckman, Cambridge, Cambridge University Press, pp. 152-178.

Beckman M. et Edwards J., 1992, Intonational categories and the articulatory control of duration, in *Speech perception, production and linguistic structure*, Y. Tohkura, E. Vatikiotis-Bateson et Y. Sagisaka, Tokyo, Ohmsha.

Beckman M. et Elam G., 1997, *Guidelines for ToBI labeling, version 3.0*, The Ohio State University Research Foundation.

Beckman M. et Pierrehumbert J., 1986, Intonational structure in Japanese and English, *Phonology Yearbook*, 3, pp. 255-309.

Beckman M. et Pierrehumbert J., 1992, Comments on chapters 14 "Downstep in Dutch: implications for a model" by van den Berg et al., and 15 "Modeling syntactic effects on downstep in Japanese" by Kubozono, in *Papers in laboratory phonology II: Gesture, segment, prosody*, R. Ladd, Cambridge U.K., Cambridge University Press, pp. 387-397.

Beddor P. S. et Hawkins S., 1990, The influence of spectral prominence on perceived vowel quality, *Journal of the Acoustical Society of America*, 87(6), pp. 2684-2704.

Benedict P. K., 1972, *Sino-Tibetan: A Conspectus*, Cambridge, Cambridge University Press.

Benguerel A.-P., 1973, Corrélats physiologiques de l'accent en français, *Phonetica*, 27(1), pp. 21-35.

Benveniste E., 1966, Convergences typologiques, *L'Homme* (La Haye, Mouton), VI(2), pp. 5-12. Réimprimé dans Benveniste E., 1974, *Problèmes de linguistique générale 2*, Paris, Gallimard, pp. 103-112.

Bernot D., 1980, *Le Prédicat en birman parlé*, Langues et civilisations de l'Asie du Sud-Est et du monde insulindien, n°8, Paris, Société d'Etudes Linguistiques et Anthropologiques de France.

Bickel B., 1995, Relatives à antécédent interne, nominalisation et focalisation: entre syntaxe et morphologie en bélharien, *Bulletin de la Société de Linguistique de Paris*, 90, pp. 391-427.

Bickel B., 2003, Referential density in discourse and syntactic typology, *Language*, 79(4), pp. 708-736.

Björverud S., 1998, *A Grammar of Lalo*, Lund, Lund University.

Bladon A. et Fant G., 1978, A two-formant model and the cardinal vowels, *STL-QPSR*, 1, pp. 1-8.

Blankenship B., 2002, The timing of nonmodal phonation in vowels, *Journal of Phonetics*, 30, pp. 163-191.

Bloch B., 1941, Phonemic overlapping, *American Speech*, 16(4), pp. 278-284.

Blomgren M., Chen Yang, Manwa L. N. et Gilbert H. R., 1998, Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers, *Journal of the Acoustical Society of America*, 103(5), pp. 2649-2658.

Bloomfield L., 1933, *Language*, New York, H. Holt and Company.

Bøckman H., 1988, *Naxi studies in China. A research report*, Oslo, University of Oslo.

Boersma P., 2001, Praat, a system for doing phonetics by computer, *Glot International*, 5(9-10), pp. 341-345.

Bolinger D. L. M., 1958, A Theory of Pitch Accent in English, *Word*, 14, pp. 109-149.

Bolinger D. L. M., 1965, *Forms of English: accent, morphemes, order*, Cambridge, Massachusetts & Tokyo, Harvard University Press/ Hokuou Publishing Company.

Bonin C.-E., 1899, Note sur un manuscrit mosso, Actes du IIe Congrès international des orientalistes [1897], Paris: E. Leroux, pp. 1-10.

Bonin C.-E., 1903, Vocabulaires, *T'oung Pao*, series 2, vol. 4, pp. 117-126.

Bouzon C., 2004, *Rythme et structuration prosodique en anglais britannique contemporain*, thèse de doctorat, Aix en Provence, Université Aix-Marseille 1.

Bouzon C. et Hirst D., 2004, Ambisyllabicité en anglais britannique contemporain, 12ème colloque de l'ALOES, Villetaneuse.

Bradley D., 1975, Nahsi and Proto-Burmese-Lolo, *Linguistics of the Tibeto-Burman Area*, 2(1), pp. 93-150.

Bradley D., 1979, *Proto-Loloish*, Scandinavian Institute of Asian Studies, Monograph series n°39, London & Malmö, Curzon Press.

Bradley D., 1985, Arakanese vowels, in *Linguistics of the Sino-Tibetan area: The state of the art*, G. Thurgood, J. A. Matisoff et D. Bradley, Canberra, Pacific Linguistics C-87, pp. 180-200.

Bradley D., 1989, Nasality as a prosody in Loloish, in *Prosodic Analysis and Asian Linguistics: to honour R.K. Sprigg*, D. Bradley, E. J. A. Henderson et M. Mazaudon, Canberra, Pacific Linguistics C-104, pp. 143-155.

Bradley D., 1994, *A Dictionary of the Northern Dialect of Lisu*, Canberra, Pacific Linguistics C-126.

Bradley D., 1995, Grammaticalisation of extent in Mran-Ni, *Linguistics of the Tibeto-Burman Area*, 18(1), pp. 1-28.

Brenzinger M., Dwyer A. M., de Graaf T., Grinevald C., Krauss M., Miyaoka O., Ostler N., Sakiyama O., Villalon M. E., Yamamoto A. Y. et Zepeda O., 2003, *Language Vitality and Endangerment; reference document for International Expert Meeting on UNESCO Program for the Safeguarding of Endangered Languages*, Paris, UNESCO.

Bruce G., 1977, Swedish word accents in sentence perspective, *Travaux de l'Institut de Linguistique de Lund*, XII.

Bruce G. et Gårding E., 1978, A prosodic typology for Swedish dialects, in *Nordic Prosody*, E. Gårding, G. Bruce et R. Bannert, Lund, Department of Linguistics, Lund University, pp. 219-228.

Burling R., 1967, *Proto-Lolo-Burmese*, Bloomington/The Hague, Mouton & Co.

Cai Hua, 1997, *Une société sans père ni mari: les Na de Chine*, collection Ethnologies, Paris, Presses Universitaires de France.

Campbell W. N., 1995, Loudness, spectral tilt, and perceived prominence in dialogues, International Congress of Phonetic Sciences, Stockholm, pp. 676-679.

Cappeau P., 2005, Quand on ne trouve pas ce que l'on cherche, *Faits de langues*, 25, pp. 157-160.

Carlson R. et Granström B., 1973, Word accent, emphatic stress, and syntax in a synthesis by rule scheme for Swedish, *Speech Technology Laboratory Quartely Progress and Status Report*, 2-3.

Carton F., 1974, *Introduction à la phonétique du français*, Paris, Bordas.

Carton F., Hirst D., Marchal A. et Séguinot A., 1976, *L'accent d'insistance*, Studia Phonetica n°12, Montréal, Didier.

Castellengo M., Chuberre B. et Henrich N., 2004, Is Voix Mixte, the vocal technique used to smoothe the transition across the two main laryngeal mechanisms, an independent mechanism?, International Symposium on Musical Acoustics (ISMA 2004), Nara, Japan.

Catford J. C., 1964, Phonation types: The classification of some laryngeal components of speech production, in *In Honour of Daniel Jones*, D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott et J. L. M. Trim, London, Longmans, Green and Co., pp. 26-37.

Catford J. C., 1977, *Fundamental Problems in Phonetics*, Edinburgh, Edinburgh University Press.

Chan M. K. M., 1993, Review of *The Prosody of Mandarin Chinese* by Xiao-nan Susan Shen, *Journal of Phonetics*, 1993(21), pp. 343-347.

Chang Nien-Chuang, 1958, Tones and Intonation in the Chengtu Dialect (Szechuan, China), *Phonetica*, 2, pp. 59-85.

Chao Yuen-ren, 1933, Tone and intonation in Chinese, *Bulletin of the Institute of History and Philology*, 4:3, pp. 121-134.

Chao Yuen-ren, 1935, Intonation of National Language [en chinois], *National Language Weekly*, 214.

Chao Yuen-ren, 1948, *Mandarin Primer*, Cambridge, Massachusetts, Harvard University Press.

Chao Yuen-ren, 1968, *A Grammar of Spoken Chinese*, Berkeley/Los Angeles, University of California Press.

Chen M. Y., 1975, An areal study of nasalization in Chinese, in *Nasálfest: Papers from a Symposium on Nasals and Nasalization*, C. A. Ferguson, L. M. Hyman et J. Ohala, Stanford, Stanford University, pp. 81-124.

Chen M. Y., 2000, *Tone Sandhi : Patterns Across Chinese Dialects*, Cambridge Studies in Linguistics 92, Cambridge, Cambridge University Press.

Chen Shilin, Bian Shiming et Li Xiuqing, 1985, *Yiyu Jianzhi [Présentation de la langue yi]*, Beijing, Minzu Chubanshe. (Voir références en chinois en fin de bibliographie.)

Chen S. H., 2005, The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects, *Journal of the Acoustical Society of America*, 117(5), pp. 3225-3230.

Chen Y., Robb M. P. et Gilbert H. R., 2002, Electroglottographic evaluation of gender and vowel effects during modal and vocal fry phonation, *Journal of Speech, Language and Hearing research*, 45, pp. 821-829.

Childers D. G., 1978, *Modern spectrum analysis*, IEEE Press.

Childers D. G. et Lee C. K., 1991, Vocal quality factors: Analysis, synthesis and perception, *Journal of the Acoustical Society of America*, 90(5), pp. 2394-2410.

Cho T., 2005, Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ, i/ in English, *Journal of the Acoustical Society of America*, 117(6), pp. 3867-3878.

Cho T. et Ladefoged P., 1999, Variation and universals in VOT: evidence from 18 languages, *Journal of Phonetics*, 27, pp. 207-229.

Chomsky N., 1972, Deep structure, surface structure and semantic interpretation, in *Studies on semantics in generative grammar*, The Hague, Mouton, pp. 62-119.

Chomsky N. et Halle M., 1968, *The Sound Pattern of English*, New York, Harper & Row.

Chuberre B., 2000, *Les registres et passages dans la voix chantée*, thèse de doctorat, Faculté de Médecine, Nantes, Université de Nantes.

Chuquet H. et Paillard M., 1989, *Approche linguistique des problèmes de traduction anglais-français*, Gap, Ophrys.

Clements N., 1979, The Description of Terraced-level Tone Languages, *Language*, 55(3), pp. 536-558.

Clements N., 1983, The hierarchical representation of tone features, in *Current Approaches to African Linguistics*, I. R. Dihoff, Dordrecht, Foris, 1, pp. 145-176.

Clements N., 1984, Principles of Tone Assignment in Kikuyu, in *Autosegmental studies in Bantu Tone*, N. Clements et J. Goldsmith, Dordrecht, Foris, pp. 281-339.

Clements N., 2000, Phonology, in *African Languages: an Introduction*, B. Heine et D. Nurse, Cambridge, Cambridge University Press, pp. 123-160.

Clements N., 2001, Representational economy in constraint-based phonology, in *Distinctive Feature Theory*, T. A. Hall, Berlin, Mouton de Gruyter, pp. 71-146.

Clements N. et Ford K. C., 1979, Kikuyu tone shift and its synchronic consequences, *Linguistic Inquiry*, 10, pp. 179-210.

Clements N. et Goldsmith J., 1984, *Autosegmental Studies in Bantu Tone*, Publications in African Languages and Linguistics, Dordrecht (Holland)/ Cinnaminson (USA), Foris Publications.

Clements N. et Laniran Y. O., 2003, Downstep and high raising: interacting factors in Yorùbá tone production, *Journal of Phonetics*, 31, pp. 203-250.

Clumeck H., 1976, Patterns of soft palate movement in six languages, *Journal of Phonetics*, 4, pp. 337-351.

Cohen A., Collier R. et 't Hart J., 1982, Declination: construct or intrinsic feature of speech pitch?, *Phonetica*, 39, pp. 254-273.

Cohen A. et 't Hart J., 1967, On the anatomy of intonation, *Lingua*, 19, pp. 177-192.

Cohn A. C., 1990, *Phonetic and Phonological Rules of Nasalization*, Los Angeles, UCLA Working Papers in Linguistics 76.

Cole J., 2000, Commentary on Shu-Hui Peng: Integrating the phonetics and phonology of tone alignment, in *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, M. Broe et J. Pierrehumbert, Cambridge UK, Cambridge University Press.

Colton R. H., 1984, Glottal waveform variations associated with different vocal intensity levels, Symposium on Care of the Professional Voice, New York, Voice Foundation.

Colton R. H. et Conture E. G., 1990, Problems and Pitfalls of Electroglottography, *Journal of Voice*, 4(1), pp. 10-24.

Comrie B., 1981, *Language Universals and Linguistic Typology*, Oxford, Blackwell.

Connell B., 2001, Downdrift, Downstep, and Declination, Typology of African Prosodic Systems Workshop, Bielefeld University, Germany.

Constantin de Chanay H., 2005, Pour une approche multidimensionnelle de l'intensité à l'oral : l'exemple de Jean-Claude Van Damme, in *Travaux linguistiques du CERLICO*

18: Intensité, comparaison, degré (2), C. Mérillou, Rennes, Presses Universitaires de Rennes, pp. 229-255.

Conteh P., Cowper E., James D., Rice K. et Szamosi M., 1983, Current Approaches to African Linguistics, in, J. Kaye, H. Koopman, D. Sportiche et A. Dugas, Dordrecht, Foris, pp. 127-137.

Cordier H., Les Mosos, *T'oung Pao*, series 2, vol. 9, pp. 663-688.

Cotte P., 1993, La linguistique anglaise entre la tradition descriptiviste et les théories contemporaines, in *Les théories de la grammaire française en France*, M. Viel, Paris, Hachette.

Cotte P., 1998, *L'explication grammaticale de textes anglais*, Paris, Presses Universitaires de France.

Coustenoble H. et Armstrong L., 1937, *Studies in French intonation*, Cambridge, Heffer.

Creissels D. et Grégoire C., 1993, La notion de ton marqué dans l'analyse d'une opposition tonale binaire: le cas du mandingue, *Journal of African Languages and Linguistics*, 14, pp. 107-154.

Creissels D. et Robert S., 1998, Morphologie verbale et organisation discursive de l'énoncé : l'exemple du tswana et du wolof, *Faits de langues*, 11-12, pp. 161-178.

Croft W., 1990, *Typology and Universals*, Cambridge, Cambridge University Press.

Cruttenden A., 1986, *Intonation*, Cambridge Textbooks in Linguistics, Cambridge, Cambridge University Press.

Crystal D., 1969, *Prosodic Systems and Intonation in English*, Cambridge, Cambridge University Press.

Crystal D., 1985, *A dictionary of linguistics and phonetics*, The Language Library, Oxford, Blackwell.

Culioli A., 1990, *Pour une linguistique de l'Énonciation I - Opérations et représentations*, Paris, Ophrys.

Culioli A., 1999a, *Pour une linguistique de l'Énonciation II - Formalisation et opérations et repérages*, Paris, Ophrys.

Culioli A., 1999b, *Pour une linguistique de l'Énonciation III - Domaine notionnel*, Paris, Ophrys.

Dainora A., 2001a, Eliminating downstep in prosodic labeling of American English, in *Proceedings of the Workshop on Prosody in Speech Recognition and Understanding*, M. Bacchiani, J. Hirschberg, D. Litman et M. Ostendorf, Red Bank, New Jersey, International Speech Communication Association, pp. 41-46.

Dainora A., 2001b, *An empirically based probabilistic model of intonation in American English*, Ph. D., The University of Chicago.

Dauer R. M., 1983, Stress timing and syllable timing reanalysed, *Journal of Phonetics*, 11, pp. 51-62.

Davies P., Lindsey G. A., Fuller H. et Fourcin A. J., 1986, Variation of glottal open and closed phases for speakers of English, *Proceedings of the Institute of Acoustics*, 8, pp. 539-46.

de Jong K., 1995, The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation, *Journal of the Acoustical Society of America*, 97, pp. 491-504.

Delais-Roussarie E., Rialland A., Doetjes J. et Marandin J.-M., 2002, The Prosody of Post-focus sequences in French, Speech Prosody, Aix-en-Provence, pp. 239-242.

Delattre P., 1958, Les indices acoustiques de la parole, *Phonetica*, 2(108-118), pp. 226-251.

Delattre P., 1965, *Comparing the phonetic features of English, French, German and Spanish : An interim report*, Heidelberg, Julius Groos Verlag.

Delattre P., 1966a, Les dix intonations de base du français, *The French Review*, 40(1), pp. 1-14.

Delattre P., 1966b, *Studies in French and Comparative Phonetics*, The Hague/London, Mouton and co.

Dell F., 1984, L'accentuation dans les phrases en français, in *Forme sonore du langage*, F. Dell, D. Hirst et J.-R. Vergnaud, Paris, Hermann.

Dell F., 2004, On recent claims about stress and tone in Beijing Mandarin, *Cahiers de linguistique - Asie Orientale*, 33(1), pp. 33-63.

Dell F. et Elmedlaoui M., 1985, Syllabic Consonants and Syllabification in Imdlawn Tashlhiyt Berber, *Journal of African Languages and Linguistics*, 7, pp. 105-130.

Dell F. et Elmedlaoui M., 2002, *Syllables in Tashlhiyt Berber and in Moroccan Arabic*, Dordrecht, Kluwer.

Deschamps A., Fournier J.-M., Duchet J.-L. et O'Neil M., 2004, *English phonology and graphophonemics*, Paris, Ophrys.

Di Cristo A., 1975, *Soixante-dix ans de recherches en prosodie*, Aix en Provence, Editions de l'Université de Provence.

Di Cristo A., 1981, La fréquence fondamentale intrinsèque des voyelles du français, *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, 7, pp. 183-207.

Di Cristo A., 1985, *De la microprosodie à l'intonosyntaxe*, Aix en Provence, Publications de l'Université de Provence.

Di Cristo A., 1998, Intonation in French, in *Intonation systems: a survey of twenty languages*, D. Hirst et A. Di Cristo, Cambridge, Cambridge University Press, pp. 195-218.

Di Cristo A., 2004, La prosodie au carrefour de la phonétique, de la phonologie et de l'articulation formes-fonctions, *Travaux interdisciplinaires du Laboratoire Parole et Langage*, 23, pp. 67-211.

Dilley L. et Shattuck-Hufnagel S., 1996, Glottalization of word-initial vowels as a function of prosodic structure, *Journal of Phonetics*, 24, pp. 423-444.

d'Imperio M., 2000, *The Role of Perception in Defining Tonal Targets and their Alignment*, Ph. D., Department of Linguistics, Ohio State University.

d'Imperio M., Nguyen N. et Munhall K., 2003, An articulatory hypothesis for the alignment of tonal targets in Italian, XVth International Congress of Phonetic Sciences, Barcelona, pp. 253-256.

Dô Thê Dung, Trân Thien Huong et Boulakia G., 1998, Intonation in Vietnamese, in *Intonation Systems: A Survey of Twenty Languages*, D. Hirst et A. Di Cristo, Cambridge, Cambridge University Press, pp. 395-416.

Doan Thiên Thuât, 1999, *Ngu âm tiêng Viêt [Phonétique du vietnamien]*, Hanoi, Vietnam, Nhà xuât ban dai hoc quôc gia hà nôi: Hanoi National University Press.

Docherty G. J. et Ladd R., Eds, (1992), *Papers in laboratory phonology II: Gesture, segment, prosody*, Cambridge U.K., Cambridge University Press.

Donohue M., 2003, The tonal system of Skou, New Guinea, in *Cross-linguistic studies of tonal phenomena: historical developments, phonetics of tone*, S. Kaji, Tokyo, Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies, pp. 329-364.

Dromey C., Stathopoulos E. T. et Sapienza C. M., 1992, Glottal airflow and electroglottographic measures of vocal function at multiple intensities, *Journal of Voice*, 6(1), pp. 44-54.

Duanmu San, 1990, *A Formal study of syllable, tone, stress and domain in Chinese languages*, Ph. D. dissertation, Cambridge, Massachusetts, MIT.

Duanmu San, 2000, *The Phonology of Standard Chinese*, New York/Oxford, Oxford University Press.

Duanmu San, 2004, Left-headed feet and phrasal stress in Chinese, *Cahiers de linguistique - Asie Orientale*, 33(1), pp. 65-104.

Duchet J.-L. et Fryd M., 1997, *Manuel d'anglais oral pour les concours*, Paris, Didier Erudition.

Dumestre G., 1987, *Le bambara du Mali: essais de description linguistique*, thèse de doctorat, Paris, Université de la Sorbonne Nouvelle/ Institut National des Langues et Civilisations Orientales.

Durand J., 2005, Tense/lax, the vowel system of English, and phonological theory, in *Headhood, Elements, Specification and Contrastivity*, P. Carr, J. Durand et C. Ewen, Amsterdam, John Benjamins, pp. 77-98.

Dwyer D., 1978, What sort of tone language is Mende?, *Studies in African Linguistics*, 9, pp. 167-209.

Earle M. A., 1975, *An acoustic phonetic study of Northern Vietnamese tones*, Santa Barbara, Speech Communications Research Laboratory, pp. 211.

Edmondson J. A., Esling J., Harris J. G., Shaoni L. et Ziwo L., 2001, The aryepiglottic folds and voice quality in the Yi and Bai languages: Laryngoscopic case studies, *Mon-Khmer Studies*, 31, pp. 83-100.

Edmondson J. A. et Yang Quan, 1988, Word-initial preconsonants and the history of Kam-Sui resonant initials and tones, in *Comparative Kadai: Linguistic studies beyond Tai*, J. A. Edmondson et D. B. Solnit, Summer Institute of Linguistics Publications in Linguistics, 86, Dallas, Summer Institute of Linguistics and the University of Texas at Arlington, pp. 143-166.

Ellis L. et Hardcastle W. J., 2002, Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data, *Journal of Phonetics*, 30, pp. 373-396.

Epstein M., 2002, *Voice Quality and Prosody in English*, Ph.D. dissertation, Linguistics, Los Angeles, University of California.

Erickson D., 2002, Articulation of extreme formant patterns for emphasized vowels, *Phonetica*, 59, pp. 134-149.

Erickson D., Fujimura O. et Menezes C., 2002, Contrastive Emphasis: Comparison of Pitch Accents with Syllable Magnitudes, Speech Prosody, Aix-en-Provence, pp. 495-497.

Erickson D., Iwata R., Endo M. et Fujino A., 2004, Effect of tone height on jaw and tongue articulation in Mandarin Chinese, International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, International Speech Communication Association Archive, http://www.isca-speech.org/archive/tal2004, pp. 53-56.

Eriksson A., Grabe E. et Traunmüller H., 2002, Perception of Syllable Prominence by Listeners with and without Competence in the Tested Language, Speech Prosody 2002, Aix-en-Provence.

Esling J., 1984, Laryngographic study of phonation type and laryngeal configuration, *Journal of the International Phonetic Association*, 14, pp. 56-73.

Esling J. et Edmondson J. A., 2002, The laryngeal sphincter as an articulator: Tenseness, tongue root and phonation in Yi and Bai, in *Phonetics and its applications: Festschrift*

for Jens-Peter Köster on the occasion of his 60th birthday, A. Braun et H. R. Masthoff, Stuttgart, Franz Steiner Verlag.

Fabre P., 1957, Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: glottographie de haute fréquence, *Bulletin de l'Académie Nationale de Médecine*, pp. 66-69.

Fagyal Z., 1995, *Aspects phonostylistiques de la parole médiatisée lue et spontanée : âge, prestige, situation, style et rythme de parole de l'écrivain M. Duras*, thèse de doctorat, Université Paris 3.

Fang Guoyu et He Zhiwu, 1995, *Naxi Xiangxing Wenzi Pu (Dictionnaire des caractères pictographiques naxi)*, Kunming, Yunnan Renmin Chubanshe. (Voir références en chinois en fin de bibliographie.)

Fant G., 1960, *Acoustic theory of speech production, with calculations based on X-ray studies of Russian articulations*, The Hague/ Paris, Mouton.

Fant G., 1970, *Acoustic theory of speech production, with calculations based on X-ray studies of Russian articulations*, The Hague/ Paris, Mouton.

Fant G., 1971, Distinctive Features and Phonetic Dimensions, in *Applications of Linguistics*, G. E. Perren et J. L. M. Trim, Cambridge, Cambridge University Press, pp. 219-239.

Fant G., 1974, *Speech sounds and features*, Current Studies in Linguistics, Cambridge, Massachusetts, MIT Press.

Fant G., 1986, Glottal flow: models and interaction, *Journal of Phonetics*, 14, pp. 393-399.

Fant G., Hertegard S. et Kruckenberg A., 1996, Focal accent and subglottal pressure, *TMH-QPSR*, 2, pp. 29-32.

Fant G. et Kruckenberg A., 1995, The voice source in prosody, International Congress of Phonetic Sciences, Stockholm, pp. 622-625.

Fant G., Liljencrants J. et Lin Q., 1985, A four-parameter model of glottal flow, *STL-QPSR*, 85(2), pp. 1-13.

Ferlus M., 1979, Formation des registres et mutations consonantiques dans les langues mon-khmer, *Mon-Khmer Studies*, 8, pp. 1-76.

Ferlus M., 1992, Essai de phonétique historique du khmer (du milieu du premier millénaire de notre ère à l'époque actuelle), *Mon-Khmer Studies*, 21, pp. 57-89.

Ferlus M., 1996, Remarques sur le consonantisme du proto kam-sui, *Cahiers de linguistique - Asie Orientale*, 25(2), pp. 235-278.

Ferlus M., 1997, Le maleng brô et le vietnamien, *Mon-Khmer Studies*, 27, pp. 55-66.

Ferlus M., 1998a, Du chinois archaïque au chinois ancien: monosyllabisation et formation des syllabes tendu/lâche (Nouvelle théorie sur la phonétique historique du chinois),

31st International Conference on Sino-Tibetan Languages and Linguistics, Lund, Sweden.

Ferlus M., 1998b, Les systèmes de tons dans les langues viet-muong, *Diachronica*, 15(1), pp. 1-27.

Ferlus M., 2000, On a mode of borrowing from Middle Chinese into Proto Tibetan: A new look at the problem of the relatinoship between Chinese and Tibetan, 33rd International Conference on Sino-Tibetan Languages and Linguistics, Ramkhamhaeng University, Bangkok & Krabi, Thailand.

Ferlus M., 2001a, *Les hypercorrections dans le thô de Làng Lo (Nghê An, Vietnam) ou les pièges du comparatisme*, Paris (texte distribué aux Journées de Linguistique d'Asie orientale).

Ferlus M., 2001b, The Origin of Tones in Viet-Muong, texte distribué à la 11th Annual Conference of the Southeast Asian Linguistic Society XI, Bangkok, Institute of Languages and Culture, Mahidol University.

Fletcher H., 1972, *Speech and Hearing in Communication*, Huntington, New York, Robert E. Krieger Publishing Company.

Flynn C.-Y., 2003, *Intonation in Cantonese*, LINCOM studies in Asian linguistics, Munich, LINCOM.

Fónagy I., 1983, *La vive voix: essais de psycho-phonétique*, "Langages et Sociétés", Paris, Payot.

Fónagy I., 1989, On status and functions of intonation, *Acta Linguistica Hungarica*, 39(1-4), pp. 53-92.

Fónagy I., 2003, Des fonctions de l'intonation: essai de synthèse, *Flambeau, Revue annuelle de la section française, Université des Langues Étrangères de Tokyo*, 29, pp. 1-20.

Fougeron C., 1999, Prosodically conditioned articulatory variations: a review, *UCLA Working Papers in Phonetics*, 97, pp. 1-68.

Fourcin A. J., 1971, First applications of a new laryngograph, *Medical and Biological Illustration*, 21, pp. 172-182.

Fox A., 1995, Principles of intonational typology, in *Studies in General and English Phonetics. Essays in Honour of Professor J. D. O'Connor*, J. W. Lewis, London, Routledge, pp. 187-210.

François A., 2004, La réduplication en mwotlap: les paradoxes du fractionnement, *Faits de langues*, 23-24, pp. 177-194.

Frei H., 1929, *La grammaire des fautes : introduction à la linguistique fonctionnelle, assimilation et différenciation, brièveté et invariabilité, expressivité*, Bellegarde, Société anonyme des arts graphiques de France.

Fry D. B., 1955, Duration and intensity as physical correlates of linguistic stress, *Journal of the Acoustical Society of America*, 27, pp. 765-768.

Fry D. B., 1958, Experiments in the Perception of Stress, *Language and Speech*, 1, pp. 126-152.

Fry D. B., 1968, Prosodic phenomena, in *Manual of phonetics*, B. Malmberg, North Holland.

Fu Maoji, 1940, *Etude de la langue moso (dialecte de Wei-Hsi), deuxième partie: Grammaire*, University of Western China. (Voir références en chinois en fin de bibliographie.)

Fu Maoji, 1981 (vol. 1) et 1984 (vol. 2), *Naxi-zu Tuhua-wenzi "Bai bianfu qu jing ji" yanjiu (Etude du manuscrit pictographique naxi "La chauve-souris blanche en quête des textes sacrés"),* Computational Analyses of Asian and African Languages Monograph Series, no. 6 et 9, Tokyo. (Voir références en chinois en fin de bibliographie.)

Fujimura O., 1971, Remarks on stop consonants. Synthesis experiments and acoustic cues, in *Form and substance: phonetic and linguistic papers presented to Eli Fischer-Jørgensen*, L. L. Hammerich, R. Jakobson et E. Zwirner, Copenhagen, Akademisk Forlag.

Fujisaki H. et Nagashima S., 1969, A model for the synthesis of pitch contours, *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, 28, pp. 53-60.

Gandour J., Gårding E. et Lindell K., 1978, Tones in Northern Kammu: a perceptual investigation, *Acta Orientalia*, 39, pp. 181-89.

Garde P., 1968, *L'Accent*, Paris, Presses Universitaires de France.

Gårding E., 1983, A generative model of intonation, in *Prosody: Models and Measurements*, A. Cutler et R. Ladd, Heidelberg, Springer, pp. 11-25.

Gårding E., 1987, Speech Act [Statement, Question, Focus] and Tonal Pattern in Standard Chinese : Constancy and Variation, *Phonetica*, 44, pp. 13-29.

Gårding E., 1998, Intonation in Swedish, in *Intonation Systems: A Survey of Twenty Languages*, D. Hirst et A. Di Cristo, Cambridge, Cambridge University Press.

Gedney W. J., 1986, Speculations on early Tai tones, in *Contributions to Sino-Tibetan Studies*, J. McCoy et T. Light, Leiden, E.J. Brill.

Gedney W. J., 1989, *Selected papers on comparative Tai studies*, Michigan Papers on South and Southeast Asia, Ann Arbor, Michigan, The University of Michigan Publications.

Gendrot C., 2003, Rôle de la qualité de la voix dans la simulation des émotions : une étude perceptive et physiologique, *PArole*, 27, pp. 137-158.

Gendrot C. et Adda-Decker M., 2004, Analyses formantiques automatiques de voyelles orales : évidence de la réduction vocalique en langues française et allemande, MIDL: Identification des langues et des variétés dialectales par les humains et par les machines, pp. 7-12.

Gendrot C., 2005, *Aspects perceptifs, physiologiques et acoustiques de différentes catégories prosodiques en français*, Ph. D., Paris, Université de la Sorbonne Nouvelle.

Gendrot C., Henrich N., Schade G., Muller F. et Expert R., 2004, Vocal folds vibratory patterns of laryngeal mechanism M0 as investigated with high speed cinematography and electroglottography, International Conference on Voice Physiology and Biomechanics.

Genetti C., 2003, Syntax, Prosody and Typology: Evidence from Prosodic Embedding in Dolakha Newar, Prosodic Interfaces, Nantes, pp. 111-116.

Geumann A., 2001, Vocal intensity: acoustic and articulatory correlates, 4th Conference on Motor Control, Nijmegen.

Gilbert H. R., Potter C. R. et Hoodin R., 1984, Laryngograph as a measure of vocal fold contact area, *Journal of Speech and Hearing Research*, 27, pp. 173-178.

Glave R. D. et Rietveld T., 1975, Is the effort dependence of speech loudness explicable on the basis of acoustic cues?, *Journal of the Acoustical Society of America*, 58, pp. 875-879.

Gobl C., 1988, Voice source dynamics in connected speech, *STL-QPSR*, 1, pp. 123-159.

Goldsmith J., 1976, *Autosegmental phonology*, Ph. D., Cambridge, Massachusetts, M.I.T., published by Garland Publishing, New York, 1980.

Goldsmith J., 1988, Prosodic trends in Bantu, in *Autosegmental Studies on Pitch Accent*, H. van der Hulst et N. Smith, Dordrecht & Providence, Foris Publications, pp. 81-93.

Gomez-Imbert E., 2001, More on the Tone versus Pitch Accent Typology: Evidence from Barasana and Other Eastern Tukanoan Languages, in *Proceedings of the symposium "Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics"*, S. Kaji, Tokyo, Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa, pp. 369-412.

Gomez-Imbert E. et Kenstowicz M., 2000, Barasana tone and accent, *International Journal of American Linguistics*, 66, pp. 419-463.

Gordon M. et Ladefoged P., 2001, Phonation types: a cross-linguistic overview, *Journal of Phonetics*, 29, pp. 383-406.

Grabe E. et Low E. L., 2002, Acoustic correlates of rhythm class, in *Laboratory Phonology 7*, C. Gussenhoven et N. Warner, Phonology and Phonetics, Berlin & New York, Mouton de Gruyter, pp. 515-546.

Grabe E., Post B., Nolan F. et Farrar K., 2000, Pitch accent realization in four varieties of British English, *Journal of Phonetics*, 28, pp. 161-185.

Gramming P., Sundberg J., Ternstrom S., Leanderson R. et Perkins W., 1988, Relationship between changes in vocal pitch and loudness, *Journal of Voice*, 2, pp. 118-126.

Grammont M., 1933, *Traité de phonétique*, Paris, Delagrave.

Greenberg J. H., 1963, Some universals of grammar with particular reference to the order of meaningful elements, in *Universals of Language*, J. H. Greenberg, London, MIT Press, pp. 73-113.

Greenberg J. H., 1974, *Language typology: A historical and analytic overview*, The Hague, Mouton.

Grønnum N., 1991, Prosodic parameters in a variety of regional Danish standard languages, *Phonetica*, 47, pp. 188-214.

Grønnum N., 1992, Comments on chapter 14 "Downstep in Dutch: implications for a model" by Rob van den Berg, Carlos Gussenhoven, and Toni Rietveld, in *Papers in laboratory phonology II: Gesture, segment, prosody*, G. J. Docherty et R. Ladd, Cambridge U.K., Cambridge University Press, pp. 359-367.

Grønnum N., 1998a, A critical remark on D. R. Ladd's *Intonational Phonology*, *Journal of Phonetics*, 26, pp. 109–112.

Grønnum N., 1998b, Intonation in Danish, in *Intonation Systems: A Survey of Twenty Languages*, D. Hirst et A. Di Cristo, Cambridge, Cambridge University Press.

Grønnum N. et Basbøll H., 2001, Consonant Length, Stød and Morae in Standard Danish, *Phonetica*, 58, pp. 230-253.

Gsell R., 1979a, *Etudes et recherches tonales (Contribution à la typologie tonale)*, exposé de synthèse/soutenance sur travaux (non publié), Paris, Université de la Sorbonne Nouvelle.

Gsell R., 1979b, *La prosodie du thai standard: tons et accent*, rapport de recherche non publié, Institut de Phonétique, Paris, Université Paris 3.

Gsell R., 1980, Remarques sur la structure de l'espace tonal en vietnamien du sud (parler de Saïgon), *Cahier d'études vietnamiennes, Département de Langues et Civilisations de l'Asie Orientale de l'Université Paris 7*, 4.

Guo Dalie, He Jiren et Yang Fuquan, 1999, *Naxi Wenhua Daguan (Panorama de la culture naxi)*, Kunming, Yunnan Minzu Chubanshe. (Voir références en chinois en fin de bibliographie.)

Guo Dalie et He Zhiwu, 1999, *Naxizu Shi (Histoire de l'ethnie naxi)*, Chongqing, Sichuan Minzu Chubanshe. (Voir références en chinois en fin de bibliographie.)

Gussenhoven C., 1983, Focus, mode and the nucleus, *Journal of Linguistics*, 19, pp. 377-417.

Gussenhoven C., 2002, Intonation and Interpretation: Phonetics and Phonology, Speech Prosody, Aix en Provence, pp. 47-57.

Gussenhoven C., à paraître, Types of Focus in English, in *Topic and Focus: Intonation and Meaning. Theoretical and Crosslinguistic Perspectives*, D. Büring, M. Gordon et C. Lee, Dordrecht, Kluwer.

Gussenhoven C. et Rietveld T., 1988, Fundamental frequency declination in Dutch: Testing three hypotheses, *Journal of Phonetics*, 16, pp. 355-369.

Gut U., 2000, The phonetic production of emphasis by German learners of English, New Sounds 2000, Amsterdam, pp. 155-157.

Haggard M., Summerfield Q. et Roberts M., 1981, Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading F_0 cues in the voiced-voiceless distinction, *Journal of Phonetics*, 9, pp. 49-62.

Haggard M. P., Ambler S. et Callow M., 1970, Pitch as a voicing cue, *Journal of the Acoustical Society of America*, 47, pp. 613-617.

Haiman J., 1978, Conditionals are topics, *Language and Speech*, 54, pp. 564-589.

Hajek J. et Maeda S., 2000, Investigating Universals of Sound Change: the Effect of Vowel Height and Duration on the Development of Distinctive Nasalization, in *Papers in Laboratory Phonology V*, M. Broe et J. Pierrehumbert, Cambridge, Cambridge University Press, pp. 52-69.

Hallé P., 1994, Evidence for tone-specific activity of the sternohyoid muscle in Modern Standard Chinese, *Language and Speech*, 37, pp. 103-124.

Halliday M. A. K., 1967, *Intonation and Grammar in British English*, The Hague, Mouton.

Han M. S. et Kim K.-O., 1974, Phonetic variation of Vietnamese tones in disyllabic utterances, *Journal of Phonetics*, 2, pp. 223-232.

Hanson D. G., Gerratt B. R. et Berke G. S., 1990, Frequency, intensity and target matching effects on photoglottographic measures of open quotient and speed quotient, *Journal of Speech and Hearing Research*, 33, pp. 45-50.

Hardcastle W. J. et Laver J., 1997, *The Handbook of Phonetic Sciences*, Oxford, Blackwell.

Harnad S., 1987, Psychophysical and cognitive aspects of categorical perception: A critical overview, in *Categorical perception: the groundwork of cognition*, Cambridge, Cambridge University Press, pp. 1-52.

Harris J. G., 2001, States of the glottis of Thai voiceless stops and affricates, in *Essays in Tai linguistics*, K. Tingsabadh et A. S. Abramson, Bangkok, Chulalongkorn University Press, pp. 3–11.

Harris Z. S., 1951, *Methods in Structural Linguistics*, Chicago, Chicago University Press.

Hartmann-So H., 1985, Morphophonemic changes in Daai Chin, in *Southeast Asian Linguistic Studies presented to André-G. Haudricourt*, R. Suriya, D. Thomas et P. Suwilai, Bangkok, Mahidol University, pp. 178-202.

Hashimoto-Yue A. O., 1987, Tone sandhi across Chinese dialects, in *Wang Li memorial volumes: English volume*, The Chinese Language Society of Hong Kong, Hong Kong, Joint Publishing Co., pp. 445-474.

Haudricourt A.-G., 1949, L'origine des particularités de l'alphabet vietnamien, *Dân Viêt-Nam*, 3, pp. 61-68.

Haudricourt A.-G., 1950, Les consonnes préglottalisées en Indochine, *Bulletin de la Société de Linguistique de Paris*, 46(172-182).

Haudricourt A.-G., 1952, Les voyelles brèves du vietnamien, *Bulletin de la Société de Linguistique de Paris*, 48(1), pp. 90-93.

Haudricourt A.-G., 1954, De l'origine des tons en vietnamien, *Journal Asiatique*, 242, pp. 69-82.

Haudricourt A.-G., 1961, Bipartition et tripartition des systèmes de tons dans quelques langues d'Extrême-Orient, *Bulletin de la Société de Linguistique de Paris*, 56(1), pp. 163-80.

Haudricourt A.-G., 1965, Les mutations consonantiques des occlusives initiales en môn-khmer, *Bulletin de la Société de Linguistique de Paris*, 60(1), pp. 160-72.

Haudricourt A.-G., 1967, La langue lakkia, *Bulletin de la Société de Linguistique de Paris*, 62(1), pp. 165-182.

He Jiren et Jiang Zhuyi, 1985, *Naxiyu Jianzhi (Présentation de la langue naxi)*, Beijing, Minzu Chubanshe. (Voir références en chinois en fin de bibliographie.)

He Zhiwu, 1987, *Naxiyu Jichu Yufa (Eléments de grammaire naxi)*, Kunming, Yunnan Minority Publishing House. (Voir références en chinois en fin de bibliographie.)

Heldner M., 2001a, *Spectral emphasis as a perceptual cue to prominence*, TMH-QPSR, pp. 51-57.

Heldner M., 2001b, Spectral Emphasis as an Additional Source of Information in Accent Detection, Prosody 2001: International Speech Communication Association Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding, Red Bank, New Jersey, pp. 57-60.

Heldner M., 2003, On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish, *Journal of Phonetics*, 31(1), pp. 39-62.

Heldner M., Strangert E. et Deschamps T., 1999, A focus detector using overall intensity and high frequency emphasis, International Conference on Speech and Language Processing, San Francisco.

Henderson E. J. A., 1952, The main features of Cambodian pronunciation, *Bulletin of the School of Oriental and African Studies*, 14(1), pp. 149-174.

Henderson E. J. A., 1961, Tone and intonation in Western Bwe Karen, in *Burma Research Society Fiftieth Anniversary Publication*, Rangoon, 1, pp. 59-69.

Henderson E. J. A., 1967, Grammar and tone in Southeast Asian languages, *Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig*, 16, pp. 171-178.

Henderson E. J. A., 1985, Feature shuffling in Southeast Asian languages, in *Southeast Asian Linguistic Studies presented to André-G. Haudricourt*, S. Ratanakul, D. Thomas et P. Suwilai, Bangkok, Mahidol University, pp. 1-22.

Henrich N., 2001, *Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception*, thèse de doctorat en acoustique, Paris, Université Paris 6.

Henrich N., d'Alessandro C., Castellengo M. et Doval B., 2004, On the use of the derivative of electroglottographic signals for characterization of non-pathological voice phonation, *Journal of the Acoustical Society of America*, 115(3), pp. 1321-1332.

Henrich N., Gendrot C., Schade G., Muller F. et Expert R., 2004, Characterization of features observed on the derivative of EGG signal by the use of high speed cinematography, International Conference on Voice Physiology and Biomechanics, Marseille.

Henrich N., Roubeau B. et Castellengo M., 2003, On the use of electroglottography for characterisation of the laryngeal mechanisms, Stockholm Music Acoustics Conference, Sweden.

Henton C., Ladefoged P. et Maddieson I., 1992, Stops in the World's Languages, *Phonetica*, 49, pp. 65-101.

Hess W., 1983, *Pitch determination of speech signals: Algorithms and devices*, Springer series in information sciences, Berlin, New York, Tokyo, Springer.

Hirano M., Vennard W. et Ohala J., 1970, Regulation of register, pitch and intensity of voice, *Folia Phoniatrica*, 22, pp. 1-20.

Hirayama H., 1984, A genealogical classification of the tone value system of Mandarin Chinese, *Gengo Kenkyu*, 86, pp. 33-53.

Hirst D. et Di Cristo A., 1998, *Intonation Systems: A Survey of Twenty Languages*, Cambridge, Cambridge University Press.

Hla Pe et Bernot D., 1974, La poésie birmane: 1300-1971, *Cahiers de poétique comparée*, 1(2), pp. 2-24.

Hoa M., 1980, *L'Accentuation en pékinois*, thèse de doctorat d'Etat, Paris, Ecole des hautes études en sciences sociales. Une version remaniée a été publiée en 1983 aux éditions Langages croisés, Paris (ouvrage publié par l'Association des Recherches interculturelles sur l'expression gestuelle et orale).

Hollien H., 1974, On vocal registers, *Journal of Phonetics*, 2, pp. 125-143.

Hollien H., Moore P., Wendahl R. W. et Michel J. F., 1966, On the nature of vocal fry, *Journal of Speech and Hearing Research*, 9, pp. 245-247.

Holmberg E. B., Hillman R. E. et Perkell J. S., 1989, Glottal air flow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch, *Journal of Voice*, 3(4), pp. 294-305.

Holt L. L. et Kluenderb K. R., 2000, General auditory processes contribute to perceptual accommodation of coarticulation, *Phonetica*, 57, pp. 170-180.

Hombert J.-M., 1977, Consonant Types, Vowel Height and Tone in Yorùbá, *Studies in African Linguistics*, 8(2), pp. 173-190.

Hombert J.-M., 1978, Consonant types, vowel quality and tone, in *Tone : a Linguistic Survey*, V. A. Fromkin, New York, Academic Press, pp. 77-111.

Hombert J.-M., Ohala J. et Ewan W. G., 1979, Phonetic explanation for the development of tones, *Language*, 55, pp. 37-58.

Honda K., 2004, Physiological factors causing tonal characteristics of speech: from global to local prosody, Speech Prosody 2004, Nara, Japan.

Hoole P., Munhall K. et Mooshammer C., 1998, Do air-stream mechanisms influence tongue movement paths?, *Phonetica*, 55(131-146).

House A. S. et Fairbanks G., 1953, The influence of consonant environment upon the secondary acoustic characteristics of vowels, *Journal of the Acoustical Society of America*, 25, pp. 105-113.

House D., 1990, *Tonal perception in speech*, Lund, Lund University Press.

House D., 2003a, Perceiving question intonation: the role of pre-focal pause and delayed focal peak, International Congress of Phonetic Sciences, Barcelona.

House D., 2003b, Perception of tone with particular reference to temporal alignment, in *Cross-linguistic studies of tonal phenomena: historical developments, phonetics of tone*, S. Kaji, Tokyo, Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies, pp. 203-216.

Householder F. W., 1971, *Linguistic Speculations*, Cambridge, Cambridge University Press.

Howard D. M., 1995, Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers, *Journal of Voice*, 9(2), pp. 163-72.

Howard D. M., Lindsey G. A. et Allen B., 1990, Toward the quantification of vocal efficiency, *Journal of Voice*, 4(3), pp. 205-12.

Howie J. M., 1974, On the domain of tone in Mandarin, *Phonetica*, 30, pp. 129-148.

Hualde J. I., Elordieta G., Gaminde I. et Smiljanic R., 2002, From pitch accent to stress accent in Basque, in *Laboratory Phonology 7*, C. Gussenhoven et N. Warner, Phonology and Phonetics, Berlin & New York, Mouton de Gruyter, pp. 547-584.

Huffman M. K., 1987, Measures of phonation type in Hmong, *Journal of the Acoustical Society of America*, 81(2), pp. 495-504.

Huggins A. W. F., 1971a, On the perception of temporal phenomena in speech, *Journal of the Acoustical Society of America*, 51(4), pp. 1279-1290.

Huggins A. W. F., 1971b, Just Noticeable Differences for segmental duration in natural speech, *Journal of the Acoustical Society of America*, 51(4), pp. 1270-1278.

Hyman L. M., 1975, Nasal states and nasal processes, in *Nasálfest: Papers from a Symposium on Nasals and Nasalization*, C. A. Ferguson, L. M. Hyman et J. Ohala, Stanford, Stanford University, pp. 249–264.

Hyman L. M., 1977, On the nature of linguistic stress, in *Studies in Stress and Accent*, L. M. Hyman, SCOPIL 4. University of Southern California, pp. 37-82.

Hyman L. M., 1993, Register tones and tonal geometry, in *The Phonology of Tone: the Representation of Tonal Register*, H. van der Hulst et K. Snider, Linguistic Models, Berlin & New York, Mouton de Gruyter, pp. 75-108.

Hyman L. M., 2001a, Comments on "Historical tonology of Japanese dialects" by Akiko Matsumori, in *Proceedings of the symposium "Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics"*, S. Kaji, Tokyo, Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA), pp. 123-128.

Hyman L. M., 2001b, Privative Tone in Bantu, in *Proceedings of the symposium "Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics"*, S. Kaji, Tokyo, Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa, pp. 237-258.

Hyman L. M., 2001c, Tone systems, in *Language typology and language universals: An international handbook*, M. Haspelmath, E. König, W. Oesterreicher et W. Raible, Berlin & New York, Walter de Gruyter, 2, pp. 1367-1380.

Hyman L. M., 2005, Word-prosodic typology, paper presented at the 'Between Stress and Tone' Conference, Leiden.

Hyman L. M. et Mathangwane J. T., 1998, Tonal domains and depressor consonants in Ikalanga, in *The theoretical aspects of Bantu tone*, L. M. Hyman et C. W. Kisseberth, Stanford University, Center for the Study of Language and Information Publications, pp. 195-230.

Hyman L. M. et VanBik K., 2002a, Tone and stem2 formation in Hakha Lai, *Linguistics of the Tibeto-Burman Area*, 25, pp. 113-121.

Hyman L. M. et VanBik K., 2002b, Tone and syllable structure of the Hakha (Lai-Chin) noun, 28th Annual Meeting of the Berkeley Linguistics Society, Special Session on Southeast Asian Linguistics, Berkeley, Berkeley Linguistics Society.

International Phonetic Association, 1999, *Handbook of the International Phonetic Association : a guide to the use of the International Phonetic Alphabet*, Cambridge, U.K. ; New York, NY, Cambridge University Press.

Issatchenko A. V. et Schädlich H. J., 1966, Untersuchungen über die deutsche Satzintonation, *Studia Grammatica*, 7, pp. 7-64.

Ito M., Tsuchida J. et Yano M., 2001, On the effectiveness of whole spectral shape for vowel perception, *Journal of the Acoustical Society of America*, 110, pp. 1141-1149.

Iwata R., Sawashima M. et Hirose H., 1981, Laryngeal adjustments for syllable-final stops in Cantonese, *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics*, 15, pp. 45-54.

Iwata R., Sawashima M., Hirose H. et Niimi S., 1979, Laryngeal adjustments of Fukienese stops: initial plosives and final applosives, *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics*, 13, pp. 61-81.

Jackendoff R., 1972, *Semantic Interpretation in Generative Grammar*, Cambridge, Massachusetts, MIT Press.

Jackson A., 1979, *An analytical appraisal of the Na-khi ritual texts*, Religion and Society 8, The Hague, Mouton.

Jackson A., 1989, Naxi studies: past, present and future, in *Ethnicity and ethnic groups in China*, Chao Chien et N. Tapp, Hong Kong, New Asia Academic Bulletin vol. VIII.

Jacques G., 2004, *Phonologie et morphologie du japhug (rGyalrong)*, thèse de doctorat, UFR de Sciences du langage, Paris, Université Paris VII.

Jakobson R., Fant G. et Halle M., 1952, *Preliminaries to Speech Analysis*, Cambridge, Massachusetts, MIT Press.

Jin Shunde, 1996, *An acoustic study of sentence stress in Mandarin Chinese*, Ph. D., Ohio State University.

Jones D., 1917, *English Pronouncing Dictionary*, London, J.M. Dent & Sons.

Jones D., 1918, *Outline of English Phonetics*, Leipzig:B.G. Teubner Verlag / Cambridge:Heffer and sons.

Jones D., 1950, *The Phoneme, its Nature and Use*, Cambridge, Heffer.

Jun S.-A., 2005, Prosodic Typology, in *Prosodic typology: the phonology of intonation and phrasing*, S.-A. Jun, Oxford, Oxford University Press.

Kam Tak Him, 1977, Derivation by tone change in Cantonese: A preliminary survey, *Journal of Chinese Linguistics*, 5(2), pp. 186-210.

Keating P., 1990, The window model of coarticulation: Articulatory evidence, in *Papers in Laboratory Phonology I*, J. Kingston et M. Beckman, Cambridge, Cambridge University Press.

Keating P., Cho T., Fougeron C. et Hsu C.-S., 2003, Domain-initial articulatory strengthening in four languages, in *Phonetic Interpretation*, J. Local, R. Ogden et R. Temple, Papers in Laboratory Phonology 6, Cambridge, Cambridge University Press, pp. 145-163.

Kenstowicz M., 1994, *Phonology in Generative Grammar*, Blackwell Textbooks in Linguistics, Cambridge, Massachusetts & Oxford, U.K., Blackwell.

Kent R.D. et Read C., 1992, *The Acoustic Analysis of Speech*, Whurr Publishers.

Kingdon R., 1958, *The Groundwork of English Intonation*, London, Longman.

Kingston J. et Beckman M., Eds, (1990), *Papers in laboratory phonology I*, Cambridge, Cambridge University Press.

Kisseberth C. W., 2001, Bantu Perspectives on Yukihiko Nakai's "The History of the Kyoto Accent", in *Proceedings of the symposium "Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics"*, S. Kaji, Tokyo, Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA), pp. 145-151.

Klatt D. H. et Cooper W., 1975, Perception of segment duration in sentence contexts, in *Structure and Process in Speech Perception*, A. Cohen et S. G. Nooteboom, Heidelberg, Springer, pp. 69-89.

Kochanski G. P. et Shih Chilin, 2001, Hierarchical structure and word strength prediction of Mandarin prosody, 4th International Speech Communication Association Tutorial and Research Workshop on Speech Synthesis, Scotland.

Kochanski G. P., Grabe E., Coleman J. et Rosner B., 2005, Loudness predicts prominence: Fundamental frequency lends little, *Journal of the Acoustical Society of America*, 118(2), pp. 1038-1054.

Kohler K. J., 1979, Dimensions in the Perception of Fortis and Lenis Plosives, *Phonetica*, 36, pp. 332-343.

Kohler K. J., 1984, Phonetic explanation in phonology: the feature fortis/lenis, *Phonetica*, 41(3), pp. 150-174.

Kohler K. J., 1985, F_0 in the production of lenis and fortis plosives, *Phonetica*, 39, pp. 199-218.

Kohler K. J., 1987, Categorical pitch perception, International Congress of Phonetic Sciences, Tallinn.

Kohler K. J., 1990, Segmental reduction in German: phonological facts and phonetic explanations, in *Speech production and speech modeling*, W. J. Hardcastle et A. Marchal, Dordrecht, Kluwer, pp. 62-92.

Kohler K. J., 1991a, Form and function of intonation peaks in German : a research project, in *Studies in German Intonation, Arbeitsberiche des Instituts für Phonetik der Universität Kiel (AIPUK) n°25*, Arbeitsberiche des Instituts für Phonetik der Universität Kiel (AIPUK), University of Kiel (Germany).

Kohler K. J., 1991b, A model of German intonation, in *Studies in German Intonation, Arbeitsberiche des Instituts für Phonetik der Universität Kiel (AIPUK) n°25*, Kiel, University of Kiel (Germany), 25.

Kohler K. J., 1991c, Terminal intonation patterns in single accent utterances of German: phonetics, phonology, and semantics, in *Studies in German Intonation*, Arbeitsberiche des Instituts für Phonetik der Universität Kiel (AIPUK), nr 25, University of Kiel (Germany).

Kohler K. J., 1992, Prosodisches Transkriptionssystem für die Etikettierung von Sprachsignaldaten, in *Arbeitsberiche des Instituts für Phonetik der Universität Kiel (AIPUK), nr 26*, Kiel, pp. 238-252.

Kohler K. J., 1995, ToBIG and PROLAB: Two prosodic transcription systems for German compared, Report at the Workshop on Prosodic Labelling, International Conference of Phonetic Sciences, Stockholm. Document disponible à l'adresse http://www.ipds.uni-kiel.de/kjk/pub_exx/kk1995_1/Prsyhand.pdf

Kohler K. J., 2000, Investigating unscripted speech: Implications for phonetics and phonology, *Phonetica*, 57, pp. 85-94.

Kohler K. J., 2001, The investigation of connected speech processes. Theory, method, hypotheses and empirical data, *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)*, 35, pp. 1-32.

Kohler K. J., 2003, Neglected Categories in the Modelling of Prosody: Pitch Timing and Non-Pitch Accents, International Congress of Phonetic Sciences, Barcelona, pp. 2925-2928.

Kohler K. J., 2004, Categorical speech perception revisited, From Sound to Sense, MIT.

Kou Bangping, 1995, *Naxizu Minjian Gequ Jicheng [A Compilation of Naxi Folk Songs]*, Kunming, Yunnan Minzu Chubanshe.

Krakow R. A., 1999, Physiological organization of syllables: a review, *Journal of Phonetics*, 27, pp. 23-54.

Kroos C., Geumann A. et Hoole P., 1999, Tongue-jaw trade-offs and naturally occuring perturbation, *Journal of the Acoustical Society of America*, 105(2), pp. 1355sq.

Kubozono H., 1992, Modeling syntactic effects on downstep in Japanese, in *Papers in laboratory phonology II: Gesture, segment, prosody*, G. J. Docherty et R. Ladd, Cambridge U.K., Cambridge University Press, pp. 368-387.

Kubozono H., 2001, Comments on "Privative Tone in Bantu" by Larry Hyman, in *Proceedings of the symposium "Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics"*, S. Kaji, Tokyo, Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa, pp. 259-265.

Kühnert B., 1996, *Die alveolar-velare Assimilation bei Sprechern des Deutschen und des Englischen - kinematische und perzeptive Grundlagen*, München, Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM), 34, pp. 175-335.

Kühnert B. et Hoole P., 2004, Speaker-specific kinematic properties of alveolar reductions in English and German, *Clinical Linguistics and Phonetics*, 18(6-8), pp. 559-575.

Kvavik K. H. et Olsen C. L., 1974, Theories and methods in Spanish intonation studies, *Phonetica*, 30, pp. 65-100.

Lacheret-Dujour A. et Beaugendre F., 1999, *La prosodie du français*, Paris, CNRS Editions.

Ladd R., 1983a, Peak features and overall slope, in *Prosody: Models and Measurements*, A. Cutler et R. Ladd, Berlin, Heidelberg, New York, Tokyo, Springer, pp. 39-52.

Ladd R., 1983b, Phonological features of intonational peaks, *Language*, 59(4), pp. 721-759.

Ladd R., 1984, Declination: a review and some hypotheses, *Phonology Yearbook*, 1, pp. 53-74.

Ladd R., 1992, An introduction to intonational phonology, in *Papers in laboratory phonology II: Gesture, segment, prosody*, G. J. Docherty et R. Ladd, Cambridge U.K., Cambridge University Press.

Ladd R., 1993, In defense of a Metrical Theory of Intonational Downstep, in *The Phonology of Tone: The Representation of Tonal Register*, H. van der Hulst et K. Snider, Berlin & New York, Mouton/ De Gruyter, pp. 109-132.

Ladd R., 1994, Constraints on the gradient variability of pitch range (or) Pitch level 4 lives!, in *Papers in Laboratory Phonology III, Phonological Structure and Phonetic Forms*, P. Keating, Cambridge, Cambridge University Press, pp. 43-63.

Ladd R., 1996, *Intonational Phonology*, Cambridge, Cambridge University Press.

Ladd R., 1998, A critical remark on *Intonational Phonology*: response to Nina Grønnum, *Journal of Phonetics*, 26, pp. 113–114.

Ladd R., Faulkner D., Faulkner H. et Schepman A., 1999, Constant "segmental anchoring" of F_0 movements under change in speech rate, *Journal of the Acoustical Society of America*, 106, pp. 1543-1554.

Ladd R., Mennen I. et Schepman A., 2000, Phonological conditioning of peak alignment of rising pitch accents in Dutch, *Journal of the Acoustical Society of America*, 107, pp. 2685-2696.

Ladd R. et Silverman K., 1984, Vowel intrinsic pitch in connected speech, *Phonetica*, 41, pp. 31-40.

Ladefoged P., 1962, *Elements of Acoustic Phonetics*, Chicago, University of Chicago Press.

Ladefoged P., 1975, *A Course in Phonetics*, Harcourt Brace College Publishers.

Laks B. et Plénat M., Eds, (1993), *De natura sonorum*, Presses Universitaires de Vincennes.

Lambrecht K., 1994, *Information structure and sentence form*, Cambridge, Cambridge University Press.

Laniran Y. O., 1992, *Intonation in Tone Languages: the Phonetic Implementation of Tones in Yorùbá*, Ph. D., Ithaca, New York, Cornell University.

Lass R., 1976, *English Phonology and Phonological Theory*, Cambridge, Cambridge University Press.

Launey M., 1994, *Une grammaire omniprédicative: essai sur la morphosyntaxe du nahuatl classique*, Sciences du Langage, Paris, CNRS Editions.

Laver J., 1980, *The phonetic description of voice quality*, Cambridge, Cambridge University Press.

Laver J., 1994, *Principles of Phonetics*, Cambridge, Cambridge University Press.

Lazard G., 1994, *L'actance*, Paris, Presses Universitaires de France.

Lea W. A., 1973, Segmental and Suprasegmental Influences on Fundamental Frequency Contours, in *Consonant Types and Tones. Southern California Occasional Papers in Linguistics 1*, L. M. Hyman, pp. 15-70.

Leben W. R., 1973, *Suprasegmental phonology*, Ph. D., Cambridge, Massachusetts, MIT.

Leben W. R., 1976, The tones in English intonation, *Linguistic Analysis*, 2, pp. 69-107.

Leben W. R., 1978, The representation of tone, in *Tone: A Linguistic Survey*, V. A. Fromkin, New York/ San Francisco/ London, Academic Press.

Lee Wai-Sum et Eric Zee, 2003, Illustrations of the IPA: Standard Chinese (Beijing), *Journal of the International Phonetic Association*, 33(109-112).

Leeden A. C. v. d., 1993, *Ma'ya: A language study: A Phonology*, Jakarta, Lembaga Ilmu Pengetahuan Indonesia & Rijksuniversiteit Leiden.

Lehiste I., 1970, *Suprasegmentals*, Cambridge, Massachusetts & London, U.K., M.I.T. Press.

Lehiste I., 1975, The phonetic structure of paragraphs, in *Structure and Process in Speech Perception*, A. Cohen et S. G. Noteboom, Berlin, Springer, pp. 195-206.

Lehiste I., 1979, Perception of sentence and paragraph boundaries, in *Frontiers of Speech Communication Research*, B. Lindblom et S. Öhman, London, Academic Press, pp. 191-201.

Lehiste I., 1982, Some phonetic characteristics of discourse, *Studia Linguistica*, 36(2), pp. 117-130.

Lehiste I. et Peterson G. E., 1961, Some basic considerations in the analysis of intonation, *Journal of the Acoustical Society of America*, 33, pp. 419-423.

Léon P., Schogt H. et Burstynsky E., 1977, *La phonologie: les écoles et les théories*, Initiation à la linguistique, Paris, Klincksieck.

Li Lincan, Zhang Kun et He Cai, 1953, *Mosuo Xiangxing Wenzi zidian [Dictionnaire des pictographes mosuo (naxi)]*, Hong Kong, Shuowenshe. (Voir références en chinois en fin de bibliographie.)

Li Rong 1998 : voir références en chinois en fin de bibliographie.

Liberman M., 1975, *The intonational system of English*, Ph. D., Cambridge, Massachusetts, MIT. Distributed by Indiana University Linguistics Club.

Liberman M. et Prince A., 1977, On stress and linguistic rhythm, *Linguistic Inquiry*, 8, pp. 249-336.

Lieberman P., 1967, *Intonation, Perception and Language*, Cambridge, Massachusetts, MIT Press.

Liénard J.-S., 1972, *Analyse, synthèse et reconnaissance de la parole*, Thèse de Doctorat d'Etat ès Sciences Appliquées, Paris, Université Paris 7.

Liénard J.-S. et Di Benedetto M. G., 1999, Effect of vocal effort on spectral properties of vowels, *Journal of the Acoustical Society of America*, 106(1), pp. 411-422.

Lindblom B., 1990, Explaining phonetic variation: a sketch of the H&H theory, in *Speech production and speech modelling*, W. J. Hardcastle et A. Marchal, Dordrecht, Kluwer, pp. 403-439.

Lisker L. et Abramson A. S., 1964, A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements, *Word*, 20, pp. 384-422.

Liu Fu, 1925, *Etude expérimentale sur les tons du chinois*, Paris, Les Belles Lettres.

Lœvenbruck H., 2000, Effets articulatoires de l'emphase contrastive sur la phrase accentuelle en français, XXIIIe Journées d'Etude sur la Parole, Aussois, pp. 165-166.

Lowe J. B. et Mazaudon M., 1994, The Reconstruction Engine: A Computer Implementation of the Comparative Method, *Association for Computational Linguistics*, 20(3), pp. 381-417.

Luksaneeyanawin S., 1983, *Intonation in Thai*, Ph. D., University of Edinburgh.

Luksaneeyanawin S., 1998, Intonation in Thai, in *Intonation Systems: A Survey of Twenty Languages*, D. Hirst et A. Di Cristo, Cambridge, Cambridge University Press.

Lyons J., 1977, *Semantics*, Cambridge, Cambridge University Press.

Maeda S., 1976, *A Characterization of American English Intonation*, Ph. D., Cambridge, Massachusetts, M.I.T.

Manuel S. Y., 1995, Speakers nasalize / ð / after / n /, but listeners still hear / ð /, *Journal of Phonetics*, 23(4), pp. 453-476.

Manuel, S.Y. 1999. "Relating language-particular coarticulation patterns to other language-particular facts." Coarticulation: Theory, data and techniques ed. by W. J. Hardcastle and N. Hewlett, 179-198. Cambridge: Cambridge University Press.

Marlett S. A., Moreno Herrera F. X. et Herrera Astorga G. G., 2005, Illustrations of the IPA: Seri, *Journal of the International Phonetic Association*, 35(1), pp. 117-121.

Martin P., 1977a, Résumé d'une théorie de l'intonation, *Bulletin de l'Institut de Phonétique de Grenoble*, 6, pp. 57-87.

Martin P., 1977b, Syntax and intonation: an integrated theory, *Toronto Semiotic Circle Monographs, Victoria University*, 2.

Martin P., 1981, Pour une théorie de l'intonation, in *L'Intonation, de l'acoustique à la sémantique*, M. Rossi, A. Di Cristo, D. Hirst, P. Martin et Y. Yishinuma, Paris, Klincksieck, pp. 234-271.

Martin P., 1996, WinPitch: un logiciel d'analyse temps réel de la fréquence fondamentale fonctionnant sous Windows, XXIV Journées d'Etude sur la Parole, Avignon, pp. 224-227.

Martin P., 2001, ToBI : l'illusion scientifique?, Actes du Colloque Journées Prosodie 2001, Grenoble, pp. 144-148.

Martinet A., 1945, *La prononciation du français contemporain, témoignages recueillis en 1941 dans un camp d'officiers prisonniers*, Paris, Droz.

Martinet A., 1955, *Economie des changements phonétiques*, Bibliotheca romanica, Series prima: manualia et commentationes, Bern, Switzerland, A. Francke AG Verlag.

Martinet A., 1956, *La Description phonologique avec application au parler franco-provençal d'Hauteville (Savoie)*, Genève, Droz.

Martinet A., 1960, *Eléments de linguistique générale*, Paris, Armand Colin.

Martinet A., 1969, *La linguistique: guide alphabétique* (ouvrage collectif dirigé par A. Martinet), Paris, Denoël.

Martinet A., 1975, *Evolution des langues et reconstruction*, Le linguiste, Paris, Presses Universitaires de France.

Martinet A., 1986, *Des steppes aux océans: l'Indo-européen et les "Indo-Européens"*, Paris, Payot.

Martinet A., 1993, *Mémoires d'un linguiste. Vivre les langues*, Paris, Quai Voltaire-Edima.

Martinet A. et Walter H., 1973, *Dictionnaire de la prononciation française dans son usage réel*, Paris/Genève, Champion/Droz.

Maspero H., 1912, Etude sur la phonétique historique de la langue annamite. Les initiales, *Bulletin de l'Ecole Française d'Extrême-Orient*, 12, pp. 1-127.

Mathieu C., 2003, *A History and Anthropological Study of the Ancient Kingdoms of the Sino-Tibetan Borderland - Naxi and Mosuo*, Mellen Studies in Anthropology, 11, Lewiston, NY, Edwin Mellen Pr.

Matisoff J. A., 1970, Glottal dissimilation and the Lahu high-rising tone: A tonogenetic case study, *Journal of the American Oriental Society*, 90, pp. 13-44.

Matisoff J. A., 1973a, *The Grammar of Lahu*, Linguistics 75, Berkeley/ Los Angeles/ London, University of California Publications.

Matisoff J. A., 1973b, Tonogenesis in Southeast Asia, in *Consonant Types and Tones*, L. M. Hyman, Southern Californial occasional papers in linguistics n°1, pp. 71-96.

Matisoff J. A., 1975, Rhinoglottophilia: the mysterious connection between nasality and glottality, in *Nasálfest: Papers from a Symposium on Nasals and Nasalization*, C. A. Ferguson, L. M. Hyman et J. Ohala, Stanford, Stanford University, pp. 265-87.

Matisoff J. A., 1999, Tibeto-Burman Tonology in an Areal Context, in *Proceedings of the symposium "Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics"*, S. Kaji, Tokyo, Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa, pp. 3-31.

Matisoff J. A., 2003, *Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction*, Berkeley & Los Angeles, University of California Press.

Matsumori A., 2001, Historical Tonology of Japanese Dialects, in *Proceedings of the symposium "Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese*

Accentology, and Other Topics", S. Kaji, Tokyo, Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa.

Mazaudon M., 1973, *Phonologie tamang: étude phonologique du dialecte tamang de Risiangku, langue tibéto-birmane du Népal*, Paris, Société d'études linguistiques et anthropologiques de France.

Mazaudon M., 1977, Tibeto-Burman tonogenetics, *Linguistics of the Tibeto-Burman Area*, 3(2).

Mazaudon M., 2003, From discourse to grammar in Tamang: topic, focus, intensifiers and subordination, in *Language Variation: Papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, D. Bradley, R. LaPolla, B. Michailovsky et G. Thurgood, Pacific Linguistics, Canberra, A.N.U., pp. 145-157.

Mazaudon M. et Lowe J. B., 1991, Du bon usage de l'informatique en linguistique historique, *Bulletin de la Société de Linguistique de Paris*, 86(1), pp. 49-87.

Mazaudon M. et Michailovsky B., 1979, *Naxi [Nahi]-English vocabulary*, Paris (non publié).

McKhann C. F., 1995, The Naxi and the Nationalities Question, in *Cultural Encounters on China's Ethnic Frontiers*, S. Harrel, Seattle, University of Washington Press, pp. 39-62.

McKhann C. F., 1998, Naxi, Rerkua, Moso, Meng: Kinship, Politics and Ritual on the Yunnan-Sichuan Frontier, in *Naxi and Moso Ethnography*, M. Oppitz et E. Hsu, Zurich, Volkerkundemuseum, pp. 23-45.

Meredith S., 1990, *Issues in the phonology of prominence*, Ph. D., Cambridge, Massachusetts, MIT.

Michailovsky B., 1975, Notes on the Kiranti Verb (East Nepal), *Linguistics of the Tibeto-Burman Area*, 2(2), pp. 183-218.

Michailovsky B., 1988a, *La langue hayu*, coll. Sciences du langage, Paris, CNRS Editions.

Michailovsky B., 1988b, Phonological typology of Nepal languages, *Linguistics of the Tibeto-Burman Area*, 11(2), pp. 25-50.

Michailovsky B., à paraître, Limbu, in *Encyclopédie des sciences du langage, Dictionnaire des langues*, D. Kouloughli et A. Peyraube, Paris, Presses Universitaires de France.

Michaud A., 2002, Conservation des langues et partage des ressources : le rôle des chercheurs dans la mise en place de banques de données, Journées d'Etude de la Parole (XXIV), Nancy, pp. 153-156.

Michaud A., 2003a, Focus et tons lexicaux: de la variété des paramètres prosodiques, in *Fonction et moyens d'expression de la focalisation à travers les langues*, A. Lacheret-Dujour et J. François, Mémoires de la Société de Linguistique de Paris, nouvelle série, tome XIII, Louvain, Peeters, pp. 13-47.

Michaud A., 2003b, The Rising tone of Naxi: from syntax and intonation to the lexicon?, Prosodic Interfaces, Nantes, pp. 149-154.

Michaud A., 2004a, Final consonants and glottalization: new perspectives from Hanoi Vietnamese, *Phonetica*, 61(2-3), pp. 119-146.

Michaud A., 2004b, A Measurement from Electroglottography: DECPA, and its Application in Prosody, Speech Prosody 2004, Nara, Japan, pp. 633-636.

Michaud A. et Vu-Ngoc T., 2004, Glottalized and Nonglottalized Tones under Emphasis: Open Quotient Curves Remain Stable, F_0 Curve is Modified, Speech Prosody 2004, Nara, Japan, pp. 745-748.

Mol H. G. et Uhlenbeck G. M., 1956, The linguistic relevance of intensity in stress, *Lingua*, 5, pp. 205-213.

Moñino Y., 2001, Le palenquero de Colombie: langue accentuelle ou langue tonale?, Conférence de l'ACBLPE sur les langues créoles, Coimbra.

Monsen R. B. et Engebretson A. M., 1983, The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction, *Journal of Speech and Hearing Research*, 26, pp. 89-97.

Montagu J., 2004, Les sons sous-jacents aux voyelles nasales en français parisien : indices perceptifs des changements, Journées d'Etude de la Parole, Fès, Maroc.

Mooshammer C. et Fuchs S., 2002, Stress distinction in German: simulating kinematic parameters of tongue-tip gesture, *Journal of Phonetics*, 30(3), pp. 337-355.

Mooshammer C., Hoole P. et Kühnert B., 1995, On loops, *Journal of Phonetics*, 23, pp. 3-21.

Morel M.-A. et Danon-Boileau L., 1998, *Grammaire de l'intonation, l'exemple du français*, Gap, Ophrys.

Morev L. N., 2001, Some occurrences of sandhi and morphonology in Tai, *Mon-Khmer Studies*, 31, pp. 101-106.

Mueggler E., 2001, *The Age of Wild Ghosts: Memory, Violence, and Place in Southwest China*, Berkeley, University of California Press.

Myers S., 2003, F_0 timing in Kinyarwanda, *Phonetica*, 60, pp. 71-97.

Nagano-Madsen Y., 1992, Mora and prosodic coordination: a phonetic study of Japanese, Eskimo and Yoruba, *Travaux de l'Institut de Linguistique de Lund*, 27.

Nagano-Madsen Y., 2003, Phonetic realization of the HL and LH accents in Japanese, in *Cross-linguistic studies of tonal phenomena: historical developments, phonetics of tone*, S. Kaji, Tokyo, Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies, pp. 243-264.

Nakai Y., 2001, The History of the Kyoto Accent, in *Proceedings of the symposium "Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics"*, S. Kaji, Tokyo, Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa, pp. 129-143.

Nespor M. et Vogel I., 1986, *Prosodic phonology*, Dordrecht, Foris.

Newman P., 1995, Hausa Tonology: Complexities in an "Easy" Tone Language, in *Handbook of Phonological Theory*, J. Goldsmith, Oxford, UK & Cambridge, Massachusetts, Blackwell, pp. 762-781.

Nguyên Nhu Y, 1999, *Dai tu diên tiêng Viêt - Comprehensive Vietnamese Dictionary*, Hanoi, Nhà Xuât ban van hoa thông tin. (En vietnamien.)

Nguyen Thi Thanh Hoa et Boulakia G., 1999, Another look at Vietnamese intonation, International Congress of Phonetic Sciences, San Francisco, pp. 2399-2402.

Nguyên Van Loi et Edmondson J. A., 1998, Tones and voice quality in modern northern Vietnamese: instrumental case studies, *Mon-Khmer Studies*, 28, pp. 1-18.

Niebuhr O., 2003, Perceptual Study of Timing Variables in F_0 Peaks, 15th International Congress of Phonetic Sciences, Barcelona, pp. 1225-2118.

Niebuhr O. et Kohler K. J., 2004, Perception and Cognitive Processing of Tonal Alignment in German, International Symposium on Tonal Aspects of Languages, With Emphasis on Tone Languages, Beijing.

Niederer B., 1998, *Les langues Hmong–Mjen (Miáo–Yáo), phonologie historique*, Lincom Studies in Asian Linguistics 7, München, Lincom Europa.

Nolan F., 1992, The descriptive role of segments: evidence from assimilation, in *Papers in laboratory phonology II: Gesture, segment, prosody*, G. J. Docherty et R. Ladd, Cambridge U.K., Cambridge University Press, pp. 261-280.

Nolan F., 1995, The effect of emphasis on declination in English intonation, in *Studies in general and English phonetics. Essays in honour of Professor J.D. O'Connor*, J. W. Lewis, London & New York, Routledge, pp. 241-254.

Nolan F., 1999, The devil is in the detail, International Congress of the Phonetic Sciences, San Francisco, pp. 1-8.

Noteboom S. G., Brokx J. P. L. et De Rooij J. J., 1978, Contributions of prosody to speech perception, in *Studies in the perception of language*, W. J. M. Levelt et G. B. Flores d'Arcais, Chichester, Wiley, pp. 75-107.

O'Connor J. D. et Arnold G. F., 1973, *Intonation of Colloquial English*, London, Longman.

Odden D., 1995, Tone: African languages, in *Handbook of Phonological Theory*, J. Goldsmith, Oxford, Blackwell.

Odden D., 1999, Typological issues in tone and stress in Bantu, Proceedings of the symposium: Cross-linguistic studies of tonal phenomena: tonogenesis, typology, and related topics, ILCAA, Tokyo University of Foreign Studies, pp. 187-215.

Ohala J., 1973, The Physiology of Tone, in *Consonant Types and Tones*, L. M. Hyman, Southern Californial occasional papers in linguistics n°1, pp. 1-14.

Ohala J., 1975, Phonetic explanations for nasal sound patterns, in *Nasálfest: Papers from a symposium on nasals and nasalization. Language Universals Project*, C. A. Ferguson, L. M. Hyman et J. Ohala, Stanford, pp. 289-316.

Ohala J., 1978, The Production of Tone, in *Tone: a linguistic survey*, V. A. Fromkin, New York/San Francisco/London, Academic Press.

Ohala J., 1990, Respiratory activity in speech, in *Speech production and speech modelling*, W. J. Hardcastle et A. Marchal, Dordrecht, Kluwer, pp. 23-53.

Ohala J. et Ewan W. G., 1973, Speed of pitch change, 84th Meeting of the Acoustical Society of America, abstract published in the Journal of the Acoustical Society of America 53(1):345.

Ohala J. et Jaeger J. J., 1986, *Experimental Phonology*, Orlando, Academic Press.

Öhman S., 1967, *Word and sentence intonation: a quantitative model*, Stockholm, Speech Transmission Laboratory Quarterly Progress and Status Report, KTH, 2-3, pp. 20-54.

Okell J. A., 1988, Notes on tone alternations in Maru verbs, in *Prosodic analysis and Asian linguistics: to honour R.K. Sprigg*, D. Bradley, E. J. A. Henderson et M. Mazaudon, Canberra, Pacific Linguistics C104, pp. 109-114.

Okrand M., 1974, Na-khi and Proto-Lolo-Burmese: A Preliminary Survey, *Linguistics of the Tibeto-Burman Area*, 1(1), pp. 55-97.

Orlikoff R. F., 1998, Scrambled EGG: The Uses and Abuses of Electroglottography, *Phonoscope*, 1(1).

Padgett J. et Tabain M., 2005, Adaptive dispersion theory and phonological vowel reduction in Russian, *Phonetica*, 62, pp. 14-54.

Palmer H. E., 1922, *English Intonation, with Systematic Exercises*, Cambridge, Heffer.

Pamies Bertrán A., 1996, Consideraciones sobre la marca acústica del acento fonológico, *Estudios de Fonética Experimental (Universidad de Barcelona)*, VIII, pp. 11-49.

Paris M.-C., 1979, *Nominalization in Mandarin Chinese: the morpheme 'de' and the 'shi...de' constructions*, Département de recherches linguistiques, Paris, Université Paris VII.

Paris M.-C., 1981, *Problèmes de syntaxe et de sémantique en linguistique chinoise*, Paris, Collège de France.

Paris M.-C., 1997, L'espace énonciatif en chinois contemporain: mais où la triade est-elle passée?, in *Diversité des langues et représentations cognitives*, C. Fuchs et S. Robert, Paris, Ophrys, pp. 93-105.

Paris M.-C., 1998, Focus operators and types of predication in Mandarin, *Cahiers de linguistique - Asie Orientale*, 27(2), pp. 139-159.

Paris M.-C., 1999, Ordre des mots, topique et focus en chinois contemporain, La thématisation dans les langues, Caen, Peter Lang, pp. 201-215.

Parker S., 2001, The acoustic qualities of Bora vowels, *Phonetica*, 58, pp. 179-195.

Passot F., 2004, *Hiérarchisation des constituants discursifs dans un corpus d'anglais oral spontané*, thèse de doctorat, Institut du Monde Anglophone, Paris, Université Paris 3 - Sorbonne Nouvelle.

Paulian C., 1975, *Le kukuya: langue teke du Congo*, Paris, Société d'Etudes Linguistiques et Anthropologiques de France.

Peng Shuhui, 1993, Cross-Language Influence on the Production of Mandarin /f/ and /x/ and Taiwanese /h/ by Native Speakers of Taiwanese Amoy, *Phonetica*, 50, pp. 245-260.

Peperkamp S. et Dupoux E., 2002, A typological study of stress 'deafness', in *Laboratory Phonology 7*, N. Warner, Phonology and Phonetics, Berlin & New York, Mouton de Gruyter, pp. 203-236.

Perkell J. S. et Nelson W. L., 1985, Variability in production of the vowels /i/ and /a/, *Journal of the Acoustical Society of America*, 77, pp. 1889-1895.

Pham A. H., 2003a, The Key Phonetic Properties of Vietnamese Tone: A Reassessment, International Congress of Phonetic Sciences, Barcelona, pp. 1703-1706.

Pham A. H., 2003b, *Vietnamese Tone: A New Analysis*, Oustanding Dissertations in Linguistics, London/New York/Oslo/Singapore, Routledge-Taylor and Francis.

Pham Thi Ngoc Yên, Castelli E. et Nguyên Quôc Cuong, 2002, Gabarits des tons vietnamiens, Journées d'Etude sur la Parole (XXIV), Nancy, pp. 25-28.

Philippson G., 1998, Evolution des systèmes prosodiques dans les langues bantu : de la typologie à la diachronie, *Faits de langues*, 11-12, pp. 429-440.

Pierrehumbert J., 1980, *The Phonology and Phonetics of English Intonation*, Ph. D. Thesis, Cambridge, Massachusetts, Massachusetts Institute of Technology (distribué par: Indiana University Linguistics Club).

Pierrehumbert J. et Beckman M., 1988, *Japanese Tone Structure*, Linguistic Inquiry Monograph 15, Cambridge, Massachusetts, MIT Press.

Pierrehumbert J. et Liberman M., 1984, Intonational invariance under changes in pitch range and length, in *Language sound structure: studies in phonology presented to Morris*

Halle by his teacher and students, R. T. Oehrle et M. Aronoff, Cambridge, Mass. & London, UK, MIT Press, pp. 157-233.

Pierrehumbert J. et Steele S., 1989, Categories of tonal alignment in English, *Phonetica*, 46, pp. 181-196.

Pike K. L., 1945, *The Intonation of American English*, University of Michigan Publications in Linguistics 1, Ann Arbor, University of Michigan Press.

Pike K. L., 1948, *Tone Languages. A Technique for Determining the Number and Type of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion*, Ann Arbor, University of Michigan Press.

Pinson T. M., 1996, *Naxi phonology - a flat phonemic statement of the Longquan dialect*, *SIL unpublished manuscript*, Lijiang/Dallas (unpublished manuscript).

Pinson T. M., 1998, *Naxi-Chinese-English Glossary, with English and Chinese Indexes*, Dallas, The Summer Institute of Linguistics.

Pisoni D. B. et Luce P. A., 1987, Trading relations, acoustic cue integration, and context effect in speech perception, in *The psychophysics of speech perception*, M. E. H. Schouten, Dordrecht, Boston, Lancaster, Martinus Nijhoff Publishers.

Pointon G. E., 1980, Is Spanish really syllable-timed?, *Journal of Phonetics*, 8, pp. 293-304.

Poser W. J., 1984, *The Phonetics and Phonology of Intonation in Japanese*, Ph. D., Cambridge, Massachusetts, MIT.

Poser W. J., 1990, Evidence for foot structure in Japanese, *Language*, 66, pp. 78-105.

Post B., 2002, French Tonal Structures, Speech Prosody 2002, Aix en Provence, pp. 583-586.

Potisuk S., Gandour J. et Harper M. P., 1996, Acoustic correlates of stress in Thai, *Phonetica*, 53, pp. 200-220.

Press W. H., Teukolsky S. A., Vetterling W. T. et Flannery B. P., 1992, *Numerical recipes in C: the art of scientific computing*, Cambridge, Cambridge University Press.

Prévost S., 1998, La notion de thème: flou terminologique et conceptuel, in *Les opérations de thématisation en français*, C. Fuchs et C. Marchello-Nizia, Montpellier, Cahiers de praxématique, 30, pp. 13-35.

Prieto P., van Santen J. et Hirschberg J., 1995, Tonal alignment patterns in Spanish, *Journal of Phonetics*, 23, pp. 429-451.

Prince A. et Smolensky P., 2004, *Optimality Theory. Constraint Interaction in Generative Grammar*, London & New York, Blackwell.

Przedlacka J., 2000, Estuary English: glottaling in the Home Counties, *Oxford University Working Papers in Linguistics, Philology and Phonetics*, 5, pp. 19-24.

Queixalós F., 1978, Etude de la parole : norme et chevauchement phonémiques en langue sikwani (guahibo), *Amerindia*, 3, pp. 115-144.

Queixalós F., 1980, *Phonologie sikuani*, thèse de doctorat en linguistique, Paris, Université Paris 4.

Remijsen B., 2002, Lexically contrastive stress accent and lexical tone in Ma'ya, in *Laboratory Phonology 7*, N. Warner, Phonology and Phonetics, Berlin & New York, Mouton de Gruyter, pp. 585-614.

Repp B. H., 1982, Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception, *Psychological Bulletin*, 92(1), pp. 81-110.

Rialland A., 1995, Review of *Autosegmental Studies on Pitch Accent*, Edited by Harry van der Hulst and Norval Smith, Dordrecht and Providence: Foris Publications, 1988, 316 pp., *Journal of Phonetics*, 23, pp. 367-372.

Rialland A., 1997, Le parcours du "downstep", ou l'évolution d'une notion, *Bulletin de la Société de Linguistique de Paris*, XCII(1), pp. 207-243.

Rialland A., 1998, Systèmes prosodiques africains: une source d'inspiration majeure pour les théories phonologiques multilinéaires, *Faits de langues*, 11-12, pp. 407-428.

Rialland A., 2001, Anticipatory Raising in Downstep Realization: Evidence for Preplanning in Tone Production, in *Proceedings of the symposium "Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics"*, S. Kaji, Tokyo, Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa, pp. 301-322.

Rialland A., 2003, Comments on Yasuko Nagano-Madsen: Phonetic realization of the HL and LH accents in Japanese, in *Cross-linguistic studies of tonal phenomena: historical developments, phonetics of tone*, S. Kaji, Tokyo, Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies, pp. 265-272.

Rialland A., 2004, A typology of question prosody in African languages, Tone and Intonation in Europe, Santorini (Greece), 9-11 September 2004, European Science Foundation.

Rialland A. et Robert S., 2001, The intonational system of Wolof, *Linguistics*, 39(5), pp. 839-939.

Rialland A. et Robert S., 2003, L'Intonation en wolof, la cohérence d'un système, in *Fonction et moyens d'expression de la focalisation à travers les langues*, A. Lacheret-Dujour et J. François, Mémoires de la Société de Linguistique de Paris, nouvelle série, tome XIII, Louvain, Peeters, pp. 161-184.

Rialland A. et Sangaré M. B., 1989, Réanalyse des tons du bambara: des tons du nom à l'organisation générale du système, *Studies in African Linguistics*, 20(1), pp. 1-27.

Ridouane R., 2004, Les mots sourds en berbère chleuh : analyses fibroscopiques et photoglottographiques, Journées d'Etude de la Parole, Fès, Morocco.

Rietveld T. et Vermillion P., 2003, Cues for perceived pitch register, *Phonetica*, 60, pp. 261-272.

Robert S., 1986, Le wolof, un exemple d'expression morphologique de l'emphase, *Bulletin de la Société de Linguistique de Paris*, LXXXI, pp. 319-341.

Robert S., 2000, Le verbe wolof ou la grammaticalisation du focus. Préconstruit: topicalisation et focalisation, in *Topicalisation et focalisation dans les langues africaines*, B. Caron, Leuven, Peeters, pp. 229-267.

Rock J., 1963-1972, *A Na-Khi -- English encyclopedic dictionary*, Serie Orientale Roma, no. 28, Roma, Instituto Italiano per il Medio ed Estremo Oriente.

Romero C., 2001, *L'intensité en français contemporain: analyse sémantique et pragmatique*, thèse de doctorat en Sciences du langage, Université Paris 8.

Rossi M., 1967, L'accent, le mot et ses limites, in *Nouvelles perspectives en phonétique*, B. Malmberg, D. B. Fry et R. Lancia, Conférences et travaux de l'Institut de phonétique, volume 1, Bruxelles, Presses Universitaires de Bruxelles, pp. 81-85.

Rossi M., 1971, L'intensité spécifique des voyelles, *Phonetica*, 24, pp. 129-161.

Rossi M., 1977, L'intonation et la troisième articulation, *Bulletin de la Société de Linguistique de Paris*, LXXII(1), pp. 55-68.

Rossi M., 1979, Le français, langue sans accent?, in *L'Accent en français contemporain [Studia Phonetica 15]*, I. Fónagy et P. Léon, Montréal/Ottawa/Paris/Bruxelles, Didier, pp. 13-51.

Rossi M., 1981, Le cadre accentuel et le mot en français et en italien, in *Problèmes de prosodie, Vol. II: Expérimentations, modèles et fonctions [Studia Phonetica 18]*, P. Léon et M. Rossi, Montréal/Ottawa/Paris/Bruxelles, Didier, pp. 9-22.

Rossi M., 1995, The evolution of phonetics: A fundamental and applied science, Conference at the Stance Académique d'hommage au Professeur Max Wajskop, Brussels.

Rossi M., 1997, Is syntactic structure prosodically recoverable?, Actes de la conférence Eurospeech, European Speech Communication Association, Rhodes, vol. I, pp. 1-8 (conférence invitée).

Rossi M., 1999, *L'Intonation, le système du français: description et modélisation*, Gap/Paris, Ophrys.

Rossi M. et Chafcouloff M., 1972, Les niveaux intonatifs, *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, 1, pp. 167-176.

Rossi M. et Di Cristo A., 1980, Un modèle de détection automatique des frontières intonatives et syntaxiques, Journées d'Etude sur la Parole, Strasbourg, pp. 213-238.

Rothenberg M., 1977, Measurement of airflow in speech, *Journal of speech and hearing research*, 20, pp. 155-176.

Rothenberg M., 1992, A multichannel electroglottograph, *Journal of Voice*, 6(1), pp. 36-43.

Rothenberg M. et Mahshie J. J., 1988, Monitoring vocal fold abduction through vocal fold contact area, *Journal of Speech and Hearing Research*, 31, pp. 338-51.

Roubeau B., Chevrie-Muller C. et Arabia-Guidet C., 1987, Electroglottographic study of the changes of voice registers, *Folia Phoniatrica*, 39, pp. 280-89.

Roubeau B., Chevrie-Muller C. et Arabia-Guidet C., 1991, Control of laryngeal vibration in register change, in *Vocal Fold Physiology: Acoustic, perceptual, and physiological aspects of voice mechanisms*, J. Gauffin et B. Hammarberg, San Diego, Singular Publishing Group, pp. 279-86.

Roubeau B., 1993, *Mécanismes vibratoires laryngés et contrôle neuro-musculaire de la fréquence fondamentale*, thèse de doctorat, Orsay, Université Paris 11.

Sagart L., 1988, Glottalised Tones in China and Southeast Asia, in *Prosodic analysis and Asian linguistics: to honour R.K. Sprigg*, D. Bradley, E. J. A. Henderson et M. Mazaudon, Canberra, Pacific Linguistics C104.

Sagart L., 1993, *Les dialectes gan. Etudes sur la phonologie et le lexique d'un groupe de dialectes chinois*, Paris, Langages croisés.

Sagart L., 1999, *The roots of Old Chinese*, Current Issues in Linguistic Theory 184, Amsterdam, John Benjamins.

Sagart L. et Xu Shixuan, 2001, History through loanwords: the loan correspondences between Hani and Chinese, *Cahiers de linguistique - Asie Orientale*, 30(1), pp. 3-54.

Saussure F. de, 1916, *Cours de linguistique générale, publié par Charles Bally et Albert Séchehaye avec la collaboration d'Albert Riedlinger*, Lausanne/Paris, Payot.

Schulman R., 1989, Articulatory dynamics of loud and normal speech, *Journal of the Acoustical Society of America*, 85, pp. 295-312.

Schwartz J.-L., Boë L.-J., Vallée N. et Abry C., 1997a, The dispersion-focalization theory of vowel systems, *Journal of Phonetics*, 25, pp. 255-286.

Schwartz J.-L., Boë L.-J., Vallée N. et Abry C., 1997b, Major trends in vowel system inventories, *Journal of Phonetics*, 25, pp. 233-253.

Selkirk E., 1980, The role of prosodic categories in English word stress, *Linguistic Inquiry*, 11, pp. 563-605.

Selkirk E., 1984, *Phonology and syntax: the relation between sound and structure*, Cambridge, Massachusetts, MIT Press.

Selkirk E., 1986, On derived domains in sentence phonology, *Phonology Yearbook*, 3, pp. 371-405.

Selkirk E., 2000, The interaction of constraints on prosodic phrasing, in *Prosody: Theory and Experiment*, M. Horne, Dordrecht, Kluwer Academic Publishers, pp. 231-261.

Selmer E. W., 1928, Noen bemerkninger om den musikalske aksent i dens forhold til den sterkt og svakt skårne aksent, in *Festskrift til Rektor J. Qvigstad*, Tromsø, Tromsø Museums Skrifter, 2, pp. 250-262.

Shadle C. H., 1985, Intrinsic fundamental frequency of vowels in sentence context, *Journal of the Acoustical Society of America*, 78, pp. 1562-1567.

Shen X.-N. S., 1990, *The Prosody of Mandarin Chinese*, University of California Press.

Shih C., Kochanski G. P., Fosler-Lussier E., Chan M. et Yuan J., 2001, Implications of Prosody Modeling for Prosody Recognition, International Speech Communication Association Workshop on Prosody in Speech Recognition and Understanding, Red Bank, New Jersey.

Shih Chilin, 1997, Declination in Mandarin, European Speech Communication Association Workshop: Theory, Models and Applications, Athens, Greece, pp. 293-296.

Shih Chilin, 2000, A Declination Model of Mandarin Chinese, in *Intonation: Analysis, Modelling and Technology*, A. Botinis, Kluwer Academic Publishers, pp. 243-268.

Shih Chilin et Kochanski G. P., 2000, Chinese Tone Modeling with Stem-ML, International Conference on Spoken Language Processing 2000, Beijing.

Shimizu K., 1971, *Comparative Jukunoid: an introductory survey*, Ph. D., Department of Linguistics and Nigerian Languages, University of Ibadan, Nigeria.

Shorto H. L., 1962, *A Dictionary of modern spoken Mon*, London, Oxford University Press.

Shrivastav R. et Sapienza C. M., 2003, Objective measures of breathy voice quality obtained using an auditory model, *Journal of the Acoustical Society of America*, 114(4), pp. 2217-2224.

Silverman K. et Pierrehumbert J., 1990, The timing of prenuclear high accents in English, in *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, J. Kingston et M. Beckman, Cambridge, U.K., Cambridge University Press, pp. 71-106.

Silverman K. E. A., Beckman M., Pitrelli J., Ostendorf M., Wightman C. W., Price P., Pierrehumbert J. et Hirschberg J., 1992, ToBI: A Standard for Labeling English Prosody, Proceedings of the 1992 International Conference on Spoken Language Processing, Banff, Canada, pp. 867-870.

Simpson A. et Scheffers M., 1995, LACS: Label-Assisted Copy Synthesis, XIIth International Congress of Phonetic Sciences, pp. 346-349.

Sluijter A., Van Heuven V. J. et Pacilly J. J. A., 1997, Spectral balance as a cue in the perception of linguistic stress, *Journal of the Acoustical Society of America*, 101, pp. 503-513.

Smalley W. A., 1963, *Manual of articulatory phonetics*, New York, Practical Anthropology.

Smith B. L., 2002, Effects of speaking rate on temporal patterns of English, *Phonetica*, 59, pp. 232–244.

Solnit D. B., 1988, The position of Lakkia within Kadai, in *Comparative Kadai: Linguistic studies beyond Tai*, J. A. Edmondson et D. B. Solnit, Summer Institute of Linguistics Publications in Linguistics, 86, Dallas, Summer Institute of Linguistics and the University of Texas at Arlington, pp. 219-38.

Sorin C., 1981, Functions, roles and treatments of intensity in speech, *Journal of Phonetics*, 9, pp. 359-374.

Sprigg R. K., 1990, Tone in Tamang and Tibetan, and the advantage of keeping register-based tone systems separate from contour-based systems, *Linguistics of the Tibeto-Burman Area*, 13(1), pp. 33-56.

Stewart J. M., 1962, *An Analysis of the Structure of the Fante Verb with Special Reference to Tone and Glottalisation*, Ph. D., London, University of London.

Stewart J. M., 1993, Dschang and Ebrié as Akan-type total downstep languages, in *The Phonology of Tone: the Representation of Tonal Register*, H. van der Hulst et K. Snider, Linguistic Models, Berlin & New York, Mouton de Gruyter, pp. 185-244.

Stone M. et Vatikiotis-Bateson E., 1995, Trade-offs in tongue, jaw, and palate contributions to speech production, *Journal of Phonetics*, 23(1), pp. 81-100.

Story B. H. et Titze I. R., 2002, A preliminary study of voice quality transformation based on modifications to the neutral vocal tract area function, *Journal of Phonetics*, 30, pp. 485–509.

Straka G., 1963, La division des sons du langage en voyelles et consonnes peut-elle être justifiée ?, in *Travaux de linguistique et de littérature*, Paris, Klincksieck, 1.

Straka G., 1965, *Album phonétique*, Québec, Presses de l'Université de Laval.

Streefkerk B., 2002, *Prominence: acoustic and lexical/syntactic correlates*, Utrecht, LOT.

Strik H. et Boves L., 1992, A physiological model of intonation, AFN 1992, University of Nijmegen.

Strik H. et Boves L., 1995, Downtrend in F_0 and P_{sb}, *Journal of Phonetics*, 23, pp. 203-220.

Sun J. T.-S., 1997, The Typology of Tone in Tibetan, Chinese Languages and Linguistics IV: Typological studies of languages in China, Taibei, Taiwan, Symposium Series of the Institute of History and Philology-Academia Sinica, pp. 485-521.

Sundberg J., 1979, Maximum speed of pitch changes in singers and untrained subjects, *Journal of Phonetics*, 7, pp. 71-79.

Suwilai P., 2003, Khmu dialects: a case of register complex and tonogenesis, in *Cross-linguistic studies of tonal phenomena: historical developments, phonetics of tone*, S.

Kaji, Tokyo, Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies, pp. 13-28.

Svantesson J.-O., 2001, Tonogenesis in Southeast Asia: Mon-Khmer and beyond, in *Proceedings of the symposium "Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics"*, S. Kaji, Tokyo, Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa, pp. 45-58.

Svantesson J.-O., 2003, Comments on Suwilai Premsrirat: Khmu dialects: a case of register complex or tonogenesis, in *Cross-linguistic studies of tonal phenomena: historical developments, phonetics of tone*, S. Kaji, Tokyo, Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies, pp. 29-36.

Svantesson J.-O. et House D., 1996, Tones and non-tones in Kammu dialects, Fonetik 96, Swedish Phonetics Conference: Speech, Music and Hearing, Quarterly Progress and Status Report TMH-QPSR, Stockholm, Royal Institute of Technology, pp. 85-87.

Swerts M., 1997, Prosodic features at discourse boundaries of different strength, *Journal of the Acoustical Society of America*, 101(1), pp. 514-521.

Syrdal A. K. et Gopal H. S., 1986, A perceptual model of vowel recognition based on the auditory representation of American English vowels, *Journal of the Acoustical Society of America*, 107(6), pp. 3438-3451.

Tabain M., 2003a, Effects of prosodic boundary on /aC/ sequences: acoustic results, *Journal of the Acoustical Society of America*, 113, pp. 516-531.

Tabain M., 2003b, Effects of prosodic boundary on /aC/ sequences: articulatory results, *Journal of the Acoustical Society of America*, 113, pp. 2834-2849.

Tabain M. et Perrier P., 2005, Articulation and acoustics of /i/ at prosodic boundaries in French, *Journal of Phonetics*, 33, pp. 77-100.

Teston B. et Ghio A., 2002, Caractéristiques de la dynamique d'un pneumotachographe pour l'étude de la production de la parole: aspects acoustique et aérodynamique, Journées d'Etude de la Parole, Nancy, pp. 337-340.

‘t Hart J., 1998, Intonation in Dutch, in *Intonation Systems: A Survey of Twenty Languages*, D. Hirst et A. Di Cristo, Cambridge, Cambridge University Press.

‘t Hart J. et Cohen A., 1973, Intonation by rule: a perceptual quest, *Journal of Phonetics*, 1, pp. 309-327.

‘t Hart J., Collier R. et Cohen A., 1990, *A Perceptual Study of Intonation*, Cambridge, U.K., Cambridge University Press.

Thein Tun U., 1982, Some acoustic properties of tones in Burmese, in *Papers in Southeast Asian Linguistics n°8: Tonation*, D. Bradley, Pacific Linguistics, Series A, n°62, pp. 77-116.

Thompson L. C., 1965, *A Vietnamese Reference Grammar*, University of Washington Press.

Thongkum T. L., 1988, Phonation types in Mon-Khmer languages, in *Voice production: Mechanisms and functions*, O. Fujimura, New York, Raven Press, pp. 319-333.

Thorsen N., 1984, F_0 timing in Danish word perception, *Phonetica*, 41, pp. 17-30.

Thurgood G., 2002, Vietnamese and tonogenesis: revising the model and the analysis, *Diachronica*, 19(2), pp. 333-363.

Thurneysen R., 1946, *A Grammar of Old Irish*, Dublin, Dublin Institute for Advanced Studies.

Titze I. R., 1988, Regulation of Vocal Power and Efficiency by Subglottal Pressure and Glottal Width, in *Vocal fold physiology: voice production, mechanisms and functions*, O. Fujimura, New York, Raven Press, pp. 227-237.

Titze I. R., 1990, Interpretation of the Electroglottographic signal, *Journal of Voice*, 4(1), pp. 1-9.

Titze I. R., 1994, *Principles of voice production*, Englewood Cliffs, Prentice-Hall.

Titze I. R., 1995, Definitions and nomenclature related to voice quality, in *Vocal Fold Physiology : Voice quality control*, O. Fujimura et M. Hirano, San Diego, California, Singular Publishing Group, pp. 335-42.

Titze I. R. et Sundberg J., 1992, Vocal intensity in speakers and singers, *Journal of the Acoustical Society of America*, 91(5), pp. 2936-2946.

Toda M., Kitamura T., Honda K. et Maeda S., 2003, *MRI kansoku ni motozuku shisatsuon seiseiji no seidô keijô to sono onkyô moderu (Fricatives sibilantes : morphologie du conduit vocal acquise par IRM et modelisation acoustique)*, Kanazawa (Japan), Technical report of the Institute of Electronics, Information and Communication Engineers, SP2003-56.

Trager G. L., 1958, Paralanguage: a first approximation, *Studies in Linguistics*, 13, pp. 1-12.

Trager G. L. et Smith H. L., 1951, *An outline of English structure*, Norman, Oklahoma, Battenburg Press.

Traunmüller H. et Eriksson A., 2000, Acoustic effects of variation in vocal effort by men, women, and children, *Journal of the Acoustical Society of America*, 107, pp. 3438–3451.

Tronnier M., 1994, Tracing Nasality with the Help of the Spectrum of a Nasal Signal, Fifth Australian Conference on Speech Science and Technology, Perth, pp. 330-335.

Tronnier M., 1995, A Method for Tracing Nasality, XIIIth International Congress of Phonetic Sciences, Stockholm, pp. 452-455.

Tronnier M., 1998, Nasal Vibration Spectra in Vowels, 135th Meeting of the Acoustical Society of America and 16th International Congress on Acoustics, Seattle.

Trubetzkoy N. S., 1939 [1976], *Grundzüge der Phonologie [traduction française: Principes de phonologie]*, Prague, Travaux du cercle linguistique de Prague 7; la version à laquelle il est fait référence est la version revue et corrigée de la traduction de J. Cantineau, publiée en 1976 aux éditions Klincksieck.

Vaissière J., 1971, *Contribution à la synthèse par règles du français*, thèse de doctorat, Grenoble.

Vaissière J., 1977, Premiers essais de l'utilisation de la durée pour la segmentation en mots dans un système de reconnaissance, 8èmes Journées d'Etudes sur la Parole, Aix-en-Provence, pp. 345-352.

Vaissière J., 1980, La structuration acoustique de la phrase française, *Annali Della Scuola Normale Superiore Di Pisa*, 10(2), pp. 529-560.

Vaissière J., 1983, Language-independent prosodic features, in *Prosody: Models and Measurements*, A. Cutler et R. Ladd, Berlin, Springer Verlag, pp. 53-66.

Vaissière J., 1986, Variance and Invariance at the Word Level, in *Invariance and Variability in Speech Process*, J. S. Perkell et D. Klatt, Lawrence Erlbaum Associates, pp. 534-539.

Vaissière J., 1989, *Contribution à l'analyse des phénomènes de parole continue lue*, Habilitation à diriger des recherches, Sciences du langage, Strasbourg.

Vaissière J., 1993, *Description of fundamental frequency in read speech in the ATR 200 sentences database*, Advanced Telecommunications Research Institute International, ATR Interpreting Telecommunications Research Laboratories.

Vaissière J., 1995, Phonetic explanations for cross-linguistic similarities, *Phonetica*, 52, pp. 123-130.

Vaissière J., 1997, Langues, prosodie et syntaxe, *Traitement Automatique des Langues*, 38(1).

Vaissière J., 2001, Changements de sons et changements prosodiques: du latin au français, *PArole*, 17-18-19, pp. 53-89.

Vaissière J., 2002, Cross-linguistic prosodic transcription: French vs. English, in *Problems and methods of experimental phonetics. In honour of the 70th anniversary of Pr. L.V. Bondarko*, N. B. Volskaya, N. D. Svetozarova et P. A. Skrelin, Moscow.

Vaissière J., 2004, The Perception of Intonation, in *Handbook of Speech Perception*, D. B. Pisoni et R. E. Remez, Blackwell Textbooks in Linguistics, Oxford, U.K. & Cambridge, Massachusetts, Blackwell.

Vaissière J. et Michaud A., à paraître, Prosodic constituents in French: a data-driven approach, in *Prosody and syntax*, I. Fónagy, Y. Kawaguchi et T. Moriguchi, coll. Usage-based linguistic informatics, Amsterdam, John Benjamins.

van der Hulst H. et Snider K., 1993, *The Phonology of Tone: the Representation of Tonal Register*, Linguistic Models, Berlin & New York, Mouton de Gruyter.

Van Driem G., 1987, *A grammar of Limbu*, Berlin, Mouton de Gruyter.

Vayra M. et Fowler C. A., 1992, Declination of supralaryngeal gestures in spoken Italian, *Phonetica*, 49(1), pp. 48-60.

Viel M. et Lilly R., 1998, *La prononciation de l'anglais*, Paris, Hachette.

Vittrant A., 2004, *La modalité et ses corrélats en birman dans une perspective comparative*, thèse de doctorat, Saint-Denis, Paris 8.

Vu-Ngoc Tuân, d'Alessandro C. et Rosset S., 2002, A phonetic study of Vietnamese tones : acoustic and electroglottographic measurements, International Conference on Speech and Language Processing, Boulder, Colorado.

Vu-Ngoc Tuân, d'Alessandro C. et Michaud A., 2005, Using open quotient for the characterization of Vietnamese glottalized tones, Eurospeech-Interspeech 2005: 9th European Conference on Speech Communication and Technology, Lisboa.

Watkins J., 1999, Closed quotient of laryngeal gestures and settings in Wa, XIVth International Congress of Phonetic Sciences, University of California at Berkeley, pp. 1017-1021.

Watkins J., 2002, *The Phonetics of Wa*, Pacific Linguistics 531, Canberra, Australian National University.

Wayland R. et Jongman A., 2003, Acoustic correlates of breathy and clear vowels: the case of Khmer, *Journal of Phonetics*, 31, pp. 181-201.

Weidert A., 1987, *Tibeto-Burman tonology*, Amsterdam studies in the theory and history of linguistic science 54, Amsterdam, Benjamins.

Wells J. C., 1990, *Pronunciation Dictionary*, Harlow, England, Longman.

Welmers W. E., 1959, Tonemics, morphotonemics, and tonal morphemes, *General Linguistics*, 4, pp. 1-9.

Welmers W. E., 1973, *African language structures*, Berkeley, University of California Press.

West P., 1999, Perception of distributed coarticulatory properties of English /l/ and /r/, *Journal of Phonetics*, 27(4), pp. 405-426.

Wetzels W. L., 2002, Fieldwork and phonological theory: Comments on Demolin, Grabe & Low, Hualde et al., and Remijsen, in *Laboratory Phonology 7*, C. Gussenhoven et N. Warner, Phonology and Phonetics, Berlin & New York, Mouton de Gruyter, pp. 615-635.

Whalen D. H. et Levitt A. G., 1995, The universality of intrinsic F_0 of vowels, *Journal of Phonetics*, 23, pp. 349-366.

Whitney W. D., 1867, *Language and the study of language*, London, Trübner.

Wightman C. W., 2002, ToBI or not ToBI?, Speech Prosody 2002, Aix en Provence.

Willems L., 1986, *Robust formant analysis*, Eindhoven, IPO report 529:1-25, Institute for Perception Research.

Williamson K., 1988, Tone and accent in Ijo, in *Autosegmental Studies on Pitch Accent*, N. Smith, Dordrecht & Providence, Foris Publications, pp. 253-278.

Wood S., 1989, The precision of formant frequency measurement from spectrograms and by linear prediction, *STL-QPSR*, 1, pp. 91-93.

Wylie T., 1959, A Standard System of Tibetan Transcription, *Harvard Journal of Asiatic Studies*, 22, pp. 261-267.

Xu Lin et Zhao Yansun, 1984, *Baiyu Jianzhi (Présentation de la langue bai)*, Beijing, Renmin Chubanshe. (Voir références en chinois en fin de bibliographie.)

Xu Yi, 1997, Contextual tonal variations in Mandarin, *Journal of Phonetics*, 25, pp. 61-83.

Xu Yi, 1998, Consistency of tone-syllable alignment across different syllable structures and speaking rates, *Phonetica*, 55, pp. 179-203.

Xu Yi, 1999, Effects of tone and focus on the formation and alignment of F_0 contours, *Journal of Phonetics*, 27(1), pp. 55-106.

Xu Yi et Wang E. Q., 2001, Pitch targets and their realization: Evidence from Mandarin Chinese, *Speech Communication*, 33, pp. 319-337.

Xuan Ke, 1999, *Naxi Guyue/ Naxi Ancient Music*, Guangdong Yinxiang Chubanshe. (Disque compact vidéo d'une représentation musicale, ISRC CN-F28-99-0021-O/V.J6.)

Yang Shih-Feng, 1984, *Report on a survey of the dialects of Szechuan*, Institute of History and Philology of Academia Sinica, special publication n°82, Taipei, Taiwan. (Voir références en chinois en fin de bibliographie.)

Yao Qian, Tan Lee et Soong F. K., 2004, Use of tone information in continuous Cantonese speech recognition, Speech Prosody 2004, Nara, Japan.

Yip M., 1980, *The Tonal Phonology of Chinese*, Cambridge, Massachusetts, Indiana University Linguistics Club.

Yip M., 1992, Prosodic Morphology in 4 Chinese Dialects, *Journal of East Asian Linguistics (Kluwer, Netherlands)*, 1, pp. 1-35.

Yip M., 1993, Tonal register in East Asian languages, in *The Phonology of Tone: the Representation of Tonal Register*, H. van der Hulst et K. Snider, Linguistic Models, Berlin & New York, Mouton de Gruyter, pp. 245-268.

Yip M., 2002, *Tone*, Cambridge, U.K., Cambridge University Press.

Yu A. C. L., 2003, Some methodological issues in phonetic typology research: Cantonese contour tone revisited, 29th Annual Meeting of the Berkeley Linguistics Society.

Yuan Jiahong, Shih Chilin et Kochanski G. P., 2002, Comparison of Declarative and Interrogative Intonation in Chinese, Speech Prosody 2002, Aix-en-Provence, pp. 711-714.

Références en chinois

Les références en chinois n'ont pu être saisies dans le logiciel EndNote, qui dans la version utilisée n'offre pas le codage Unicode. Elles sont reprises ci-dessous par ordre alphabétique des auteurs dans la transcription *pinyin*. La colonne « cat. » (catégorie) indique s'il s'agit d'un livre (L) ou d'un article (A).

Auteur	Année	Titre	Editeur / Périodique	Ville	cat.
卜金荣 (主编)	1999	纳西东巴文化要籍及传承概览	云南民族出版社	昆明	L
戴庆厦	1993	关于纳西语的松紧元音问题	"民族语文"第一期, 27-31页	北京	A
方国瑜，和志武	1995	纳西象形文字谱	云南人民出版社	昆明	L
冯胜利	2000	汉语韵律句法学	上海教育出版社	上海	L
傅懋勣	1940	维西麽些语研究	西南联合大学	重庆	L
傅懋勣	vol. 1 : 1981, vol. 2 : 1984	纳西语图画文字 "白蝙蝠取经记" 研究	Computational Analyses of Asian and African Languages: Monograph Series 6	Tokyo	L
郭大烈，和志武	1999	纳西族史	四川民族出版社	重庆	L
郭大烈，郑卫东	1998	纳西族谚语-科空	云南民族出版社	昆明	L
郭大烈, 和即仁, 杨福泉	1999	纳西文化大观	云南民族出版社	昆明	L
和即仁，姜竹仪	1985	纳西语简志	民族出版社	北京	L
和志武	1987	纳西语基础语法	云南民族出版社	昆明	L
黄布凡	1991	汉藏语概论	北京大学出版社	北京	L
姜竹仪	1985	纳西语西部方言音位系统中的几个问题	"民族语文"第二期, 28-30页	北京	A
李霖灿，张琨，和才	1953	麽些象形文字字典	说文社；2001年云南民族出版社再版："纳西象形标音文字字典"	香港	L
李荣 (主编)	1998	成都方言词典	江苏教育出版社	南京	L
孙宏开	2001	纳西语在藏缅语族语言中的历史地位	"语言研究"第一期 (总第42期), 90-99页	北京	A
王洪君	1999	汉语非线性音系学	北京大学出版社	北京	L
吴洁敏，朱宏达	2001	汉语节律学	语文出版社	北京	L
徐琳，赵衍荪	1984	白语简志	民族出版社	北京	L
杨焕典	1984	论纳西语音位系统	Computational Analyses of Asian and African Languages: Monograph series 22, 131-146页	Tokyo	A
杨焕典	1994	再论关于纳西语中的紧松元音问题：兼答戴庆厦教授	Communication à la 27e Conférence internationale sur les langues et la linguistique sino-tibétaine	Paris	A
杨时逢	1984	四川方言调查报告	中央研究院历史语言研究所	台北	L
赵兴文，和民达，和元庆	1987	纳西民歌选	云南民族出版社	昆明	L
赵元任	2000	语言问题	商务印书馆	北京	L
赵元任	2002	赵元任语言学论文集	商务印书馆	北京	L

UNIVERSITE PARIS 3 – SORBONNE NOUVELLE

ECOLE DOCTORALE 268 « LANGAGE ET LANGUES »

Prosodie de langues à tons (naxi et vietnamien), prosodie de l'anglais : éclairages croisés

thèse de Doctorat nouveau régime

présentée par Pierre-Alexis MICHAUD

discipline : Phonétique

Directrice de thèse : Mme Jacqueline VAISSIÈRE

soutenue le 12 décembre 2005 devant le jury composé de :

M. Claude DELMAS, Examinateur

M. Klaus KOHLER, Prérapporteur

Mme Martine MAZAUDON, Prérapporteur

Mme Aliyah MORGENSTERN, Examinatrice

M. Mario ROSSI, Prérapporteur

Mme Jacqueline VAISSIÈRE, Directrice

Volume 2 (Annexes et figures)

Table des matières de l'Annexe 1 (« Observations sur les phonèmes du naxi »)

0.1. Brève présentation du naxi. La question de sa position vis-à-vis du groupe birman-yi 297
0.2. Variété dialectale. Parlers étudiés 298
0.3. Principaux enjeux de la description phonémique du naxi 300

1. Les initiales 305

1.1. L'initiale palatale /ç/, fruit d'anciennes initiales nasales ? 306
1.2. Allophones rétroflexes des consonnes coronales, et rétroflexion phonémique 312
1.3. Allophones palatalisés des vélaires devant /i/, /y/, /jʏ/, /ja/, /jɑ/, /jə/ 318
1.4. Autres observations concernant les initiales 324
1.4.1. Allophones « trills » des occlusives bilabiales /p^h/, /p/, /b/, /m͡b/ devant /ɣ/ 324
1.4.2. Allophones uvulaires des vélaires devant les voyelles d'arrière 325
1.4.3. Fricatives 325
1.4.4. Nasales 326
1.4.5. Latérale l, et confusions avec n͡d 327
1.4.6. Occlusives et affriquées : description en quatre degrés de *Voice Onset Time* 327

2. Observations synchroniques sur les rimes 329

2.1. Syllabes labialisées : rime /wə̹/, initiales labialisées, ou médiale /w/ ? 330
2.2. La fricative /ɣ/, noyau de syllabe 333
2.3. Notation des voyelles d'aperture moyenne 335
2.4. Réalisation phonétique des voyelles d'arrière non arrondies 336
2.5. Notation des voyelles ouvertes 337
2.6. La question de la « voyelle neutre » 337
2.7. Attaque des voyelles syllabiques 338
2.8. Possibilité de fusion entre une syllabe sans consonne initiale et la syllabe qui précède 340
2.9. Semi-voyelles 342
2.10. La voyelle rhotique : confirmation de l'absence d'oppositions de qualité de voix en naxi 344

3. Prononciation des phonèmes de la variété locale de chinois 345

3.1. Remarques générales 345
3.1.1. Situation actuelle de la langue par rapport au chinois 345
3.2. La langue chinoise parlée par l'informateur M4 348
3.2.1. Maintien intermittent d'une distance phonétique entre naxi et chinois 348
3.2.2. Suppression des nasales finales, et phénomènes d'hyper-correction 350
3.2.3. Conservation des oppositions portées par les nasales lorsqu'elles sont en position initiale 350
3.2.4. Changements dont le « chinois deuxième langue » est potentiellement porteur 351

4. Phénomènes à l'échelle du mot : fusion syllabique et intégration nominale 357

4.1. Fusion syllabique 357
4.2. Phénomènes d'intégration nominale 358
4.2.1. Modifications tonales 358
4.2.1.1. Exposition des faits 358
4.2.1.2. Hypothèses sur l'origine diachronique du changement tonal 361
4.2.2. Modifications de phonèmes 363

5. Variations intonatives dans la réalisation des phonèmes 365

6. Aperçu de la variation dialectale 367

6.1. Dialecte de FK (locuteur M12) 367
6.2. Dialecte de NL (locuteur M10) 372
6.3. Eléments de comparaison entre dialectes 373

Eléments de conclusion 378

Annexe 1. Observations sur les phonèmes du naxi

Les sections 1 et 2 présentent l'inventaire phonémique du naxi occidental : les *initiales* (section 1), et les *rimes* (section 2), notions qui en naxi se recoupent dans une large mesure avec celles de *consonnes* et *voyelles*. La section 3 est consacrée à la prononciation du chinois par les locuteurs natifs du naxi : façon d'aborder le domaine du contact entre langues sans se limiter aux seuls emprunts lexicaux. La section 4 aborde les modifications phonétiques qui ont lieu dans les mots disyllabiques ou polysyllabiques, parfois jusqu'à la fusion de deux syllabes. La section 5 décrit une partie de la variation allophonique liée à l'intonation. Enfin, la section 6 réunit des observations sur deux dialectes, et propose l'ébauche d'une comparaison.

Dans ces réflexions, qui sont principalement synchroniques, est tenté un éclairage réciproque de la synchronie et de la diachronie[1].

0.1. Brève présentation du naxi. La question de sa position vis-à-vis du groupe birman-yi (« birman-ni », « lolo-birman »)[2]

Le naxi (en chinois : na^{4} xi^{1} yu^{3} 纳西语 ; autonyme dans les dialectes occidentaux : /nù hi^{-}/) est une **langue tibéto-birmane très éloignée du proto-tibéto-birman** tel qu'il est reconstruit par Paul Benedict (Benedict 1972, *Sino-Tibetan: A Conspectus*) et James Matisoff (Matisoff 2003, *Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction*). A l'instar des langues ni/yi (en chinois : 彝语 yi^{2}yu^{3} ; langues anciennement désignées comme *lolo*), dont il est génétiquement et géographiquement proche, le naxi est une langue monosyllabique, sans flexion, dont la structure syllabique est très simple : (C)(G)V, où C représente une consonne, G (pour *Glide*) une semi-voyelle /w/ ou /j/ (dont la distribution est très limitée), et V une voyelle, seul composant indispensable de la syllabe.

---

[1] Cette perspective rejoint, nous semble-t-il, les réflexions d'André Martinet (Martinet 1955, 1975) Larry Hyman (Hyman 1975, « Nasal states and nasal processes ») et Boyd Michailovsky (voir en particulier la synthèse de travaux présentée pour l'Habilitation à diriger des recherches, Paris, 2004).

[2] La désignation *ni* pour le groupe communément appelé *yi* a été récemment proposée par David Bradley : *ni* est un autonyme relativement courant parmi les langues concernées (les Hani, les Sani…), tandis que *yi* et *lolo* sont des exonymes, le second étant de plus péjoratif (Bradley 1995:1, note 2). Tout récemment, Bradley a proposé de corriger à nouveau cette désignation, en *ngwi*. Le compromis choisi ici consiste à employer la désignation *yi*, rappelant fréquemment son équivalence avec *ni* et *ngwi* (appellations plus adéquates au plan étymologique) et *lolo* (appellation qui était la plus répandue jusqu'à une date récente).

L'affiliation précise du naxi, généralement classé à l'intérieur du sous-groupe birman-yi (« lolo-birman »), **demeure incertaine** (Bradley 1975, « Nahsi and Proto-Burmese-Lolo » ; des remarques préliminaires sur la question sont proposées par Okrand 1974). **L'analyse phonémique de dialectes naxi doit permettre de poser quelques jalons en vue d'un travail comparatif affiné.**

0.2. Variété dialectale. Parlers étudiés

Au plan historique et ethnologique, la nationalité naxi recouvre des populations qui ne se perçoivent pas nécessairement comme membres d'une même « ethnie ». Cette situation n'est pas exceptionnelle : toute division en catégories de la réalité infiniment complexe des sociétés comporte une part d'arbitraire ; le découpage en nationalités (min^{2} zu^{2} 民族) opéré après la fondation de la République populaire de Chine en 1949 a été déterminé par de multiples facteurs, dont les traditions anthroponymiques chinoises anciennes et les projets politiques et socio-économiques de la jeune République populaire de Chine.

Dans le cadre des enquêtes linguistiques nationales menées peu après 1949, l'ensemble de la région de Lijiang a fait l'objet en 1956 d'une enquête mobilisant huit linguistes ; une partie des données a plus tard été publiée sous forme de livre (He Jiren et Jiang Zhuyi 1985). L'enquête a conclu à une division du naxi en *naxi occidental* et *naxi oriental*. **Aujourd'hui, les locuteurs du naxi *oriental* se revendiquent comme « Mosuo »** (mo^{2} suo^{1} 摩梭), distincts des « Naxi ». Le terme « Mosuo » n'est pas un autonyme, mais une désignation chinoise ancienne (qui comporte plusieurs variantes, dont 麽些 mo^{1} xie^{1}), remplacée au cours du XX^{e} siècle (plus précisément, dans l'usage officiel, lors de la définition des *nationalités*, après 1949) par « Naxi » (na^{4} xi^{1} 纳西), avec la même extension. (Ainsi, le dictionnaire de pictogrammes de Li Lincan, Zhang Kun et He Cai, publié à l'origine comme « Dictionnaire des pictogrammes moxie » [Hong Kong, 1953], a été réimprimé en 2001 en Chine continentale sous le titre « Dictionnaire des pictogrammes naxi ».) L'étiquette « Mosuo », péjorative à l'origine, n'en a pas moins aujourd'hui la faveur des intéressés, qui la revendiquent haut et fort. Elle présente en effet plusieurs avantages. Au plan phonétique, « Mosuo » et « Naxi » sont bien différenciés, ce qui répond au souhait d'afficher une différence que les autonymes ne font pas ressortir : les autonymes des communautés qui se reconnaissent dans l'appellation « Mosuo » sont [nɑ̀ ˉ], [nɑ̄ z̪̄] ou [nɑ̄ hɪ̃ ˉ] selon les dialectes considérés (He Jiren et Jiang Zhuyi 1985:107), noms proches de l'autonyme des « Naxi » (prononcé [nɑ̀ hi ˉ] ou [nɑ̀ hɪ̃ ´] selon les dialectes), au point qu'ils risqueraient de jeter un doute sur la légitimité d'une séparation établie entre les deux groupes. Le terme « Mosuo » présente en outre le double avantage de se relier à une longue tradition écrite, puisqu'il était employé dans les annales chinoises depuis la dynastie Tang (唐朝, 618-907), et d'avoir disparu, au milieu du XX^{e} siècle, de l'usage officiel (remplacé par « Naxi »), ce qui

contribue à lui conférer un charme désuet et un élément de mystère, atouts de poids à une époque où l'industrie touristique (chinoise et internationale) est gourmande de couleur locale, de tradition et d'authenticité. Les « Mosuo » sont désormais reconnus, à l'échelle provinciale mais non à l'échelle nationale, comme un *peuple* (ren^{2} 人), sous-groupe à l'intérieur de la *nationalité* (min^{2} zu^{2} 民族) naxi.

Au plan historique, l'extension du naxi occidental correspond dans une large mesure à celle du domaine des seigneurs féodaux de Lijiang, les Mu (木氏土司 mu^{4} shi^{4} tu^{3} si^{1}), **vassaux de la Chine des Ming (明朝, 1368-1644) puis des Qing (清朝, 1644-1911)**, tandis que **le naxi dit *oriental*** (ou *mosuo* : mo^{2} suo^{1} yu^{3} 摩梭语) **est parlé dans une région qui échappait au contrôle des seigneurs féodaux de Lijiang, et se trouvait dans l'aire d'influence culturelle du Tibet**[1].

Des travaux ethnographiques relativement nombreux sont consacrés aux Naxi/Mosuo (les ethnologues ayant une préférence pour l'appellation *Na*), depuis les travaux pionniers mais lacunaires de Charles-Eudes Bonin, Henri Cordier et Jacques Bacot (Bonin 1899, 1903 ; Cordier 1908, « Les Mosos » ; Bacot 1913, *Les Mosso, ethnographie des Mosso, leurs religions, leur langue et leur écriture, avec les documents historiques et géographiques relatifs à Li-kiang par Ed. Chavannes*) : voir notamment Jackson 1979, Cai Hua 1997, Mathieu 2003. De grands progrès restent en revanche à réaliser dans l'histoire des langues et des populations de la région. Des aperçus des travaux menés sur la langue et la culture naxi sont proposés par Harald Bøckman (Bøckman 1988, *Naxi studies in China. A research report*) et Anthony Jackson (Jackson 1989, « Naxi studies: past, present and future »), et, en chinois, par Bu Jinrong (卜金荣, 纳西东巴文化要籍及传承概览, 昆明：云南民族出版社, 1999 ; titre anglais : *A Survey of Naxi Dongba Culture Books and Inheritance*).

Les dialectes dont il est ici question, étudiés de première main, sont les suivants :

1) parler du village de /ā ʂə̀/, hameau de /lé pȳ lò/ (en chinois : jin^{1} shan^{1} xiang^{1}, wen^{2} hua^{4} xing^{2} zheng^{4} cun^{1}, leng^{3} bu^{4} luo^{2} zi^{4} ran^{2} cun^{1} 丽江县, 金山乡, 文化行政村, 冷不罗自然村), abrégé en AS

2) parler du village de /ndā lè/ (en chinois : jin^{1} shan^{1} xiang^{1}, gui^{4} feng^{1} da^{4} lai^{2} xing^{2} zheng^{4} xia^{4} cun^{1}/da^{4} lai^{2} er^{4} cun^{1} 丽江县, 金山乡, 贵峰大来行政下村 /大来二村), abrégé en NL

[1] Voir l'histoire du peuple naxi de Guo Dalie et He Zhiwu (Guo Dalie et He Zhiwu 1999), ainsi que les réflexions de Charles McKhann concernant la division en « Mosuo » et « Naxi » (McKhann 1995, « The Naxi and the Nationalities Question » ; McKhann 1998), et les observations d'Erik Mueggler au sujet de la situation des minorités du Yunnan dans la Chine contemporaine (Mueggler 2001, *The Age of Wild Ghosts: Memory, Violence, and Place in Southwest China*).

3) parler du village de /fɣ̩ kʰō/ (en chinois : yu^4 long^2 xian^4, feng^4 ke^1 xiang^1, shan^4 mei^3 xing^2 zheng^4 cun^1, heng^2 ke^3 er^4 zu^3 玉龙县，奉科乡，善美行政村，恒可二组), abrégé en FK

Les cartes 1 et 2 indiquent la localisation de ces points d'enquête, ainsi que du parler de la ville de Lijiang elle-même, abrégé en DYZ , pour Dayanzhen (da^4 yan^2 zhen^4 大研镇), désignation du centre historique de la ville.

Les deux premiers de ces dialectes appartiennent à la sphère du naxi occidental, variété la mieux décrite (Li Lincan, Zhang Kun *et al.* 1953 ; Mazaudon et Michailovsky 1979 ; Fu Maoji 1984 ; He Jiren et Jiang Zhuyi 1985 ; Fang Guoyu et He Zhiwu 1995 ; Pinson 1998). Le dialecte de AS possède un inventaire phonémique plus étendu que le dialecte de la ville de Lijiang (DYZ) ; celui de NL se caractérise par des inversions tonales régulières par rapport aux autres dialectes (DYZ et AS). **Le troisième dialecte étudié, FK, à la frontière entre les domaines du naxi occidental et du naxi oriental** (ou *mosuo* : mo^2 suo^1 yu^3 摩梭语), ouvre des perspectives plus larges. Il est difficilement compréhensible par les locuteurs de AS, NL et DYZ, tandis que ces trois derniers dialectes sont mutuellement compréhensibles.

Le parler de Lijiang, principale ville de la région naxi, au centre du domaine du naxi occidental (abrégé en DYZ), est en fait assez peu homogène, la ville étant un point de rencontre de locuteurs de diverses régions.

Par commodité, les données citées de seconde main sont réécrites selon notre analyse phonémique lorsqu'une familiarité de première main avec le dialecte concerné démontre que la même analyse phonémique s'applique. En particulier, les tons sont réécrits comme des tons ponctuels, choix qui se fonde en partie sur la présence de tons H (Haut) flottants créateurs de contours, similaires à ceux de langues subsahariennes (tels qu'ils sont décrits par Clements et Goldsmith 1984 ; Odden 1995 ; Hyman 2001 ; les faits naxi sont présentés dans une communication de colloque : Michaud 2003, et un article soumis pour publication).

Les quatre tons des dialectes de DYZ et AS (haut H, moyen M, bas L, et montant LH) sont décrits en détail dans le corps de la thèse ; ces descriptions ne sont pas reprises dans la présente Annexe, où figurent seulement des indications sur les tons des dialectes de NL et FK (sections 6.1 et 6.2).

0.3. Principaux enjeux de la description phonémique du naxi

En synchronie, le naxi soulève des questions délicates de découpage de la syllabe en initiale et rime. Cette observation (due à Boyd Michailovsky, séminaire « Langues tibéto-birmanes ») rejoint les réflexions générales de James Matisoff sur les langues

monosyllabiques : celui-ci parle de « tightly interdependent neighbouring vowels and consonants ».

> There is something about the tightly structured nature of the syllable in monosyllabic languages which favours the shift in contrastive function from one phonological feature of the syllable to another. (Matisoff 1973b)

Des observations du même ordre sont rapportées par Eugenie Henderson (Henderson 1985, « Feature shuffling in Southeast Asian languages »). Sont abordés ici, entre autres questions, l'**opposition entre voyelles nasales /ĩ / et /ỹ / dans le dialecte de FK**, opposition qui dans les deux autres dialectes étudiés, AS et NL, subsiste partiellement sous la forme d'une opposition entre consonnes initiales fricatives ; le **découpage en initiale et rime des syllabes à consonnes vélaires palatalisées** ; et **deux oppositions présentes dans le dialecte de AS tandis qu'elles sont absentes des deux autres dialectes** : consonnes coronales rétroflexes et non rétroflexes devant /o/, syllabes avec et sans arrondissement médial devant la voyelle /ə̹/.

Le tableau 1 présente les combinaisons entre initiales et rimes attestées dans le dialecte de AS. Les combinaisons entre parenthèses sont marginales. (Les questions posées par le découpage en initiale et rime, et la possibilité d'une réanalyse reconnaissant une médiale /w/, sont abordées en détail au cours de l'analyse.)

	i	y	ɯ	u	e	ɤ	o	a	ɑ	ə˞	ɣ	wɑ	wa	wɤ	wə˞	jɤ	jɑ	ja	jə˞
Ø	ji	jy	ɣɯ	wu	(e)	ɣɤ	ø	a	(ɑ)	ə˞	ɣ	wɑ	wa	(wɤ)	(və˞)	jɤ	jɑ	ja	
pʰ	pʰi	pʰy	pʰɯ	pʰu	pʰe	pʰɤ	pʰo	pʰa	pʰɑ	pʰə˞	pʰɣ					pʰjɤ			
p	pi	py	pɯ	pu	pe	(pɤ)	po	(pa)	pɑ	pə˞	pɣ					pjɤ			
b	bi	by	bɯ	bu	be	bɤ		ba	bɑ	bə˞	bɣ								
mb	mbi	mby	mbɯ	mbu	mbe	(mbɤ)		mba	mbɑ	mbə˞	mbɣ								
m	mi	my	mɯ	mu	me	mɤ	mo	ma	mɑ	mə˞	mɣ					mjɤ		mja	
tʰ			ʈʰɯ		tʰe		tʰo	tʰa	tʰɑ	ʈʰə˞	tʰɣ								
ʈʰ							ʈʰo												
t	ti	ty	ʈɯ		te		to	ta	tɑ	ʈə˞	tɣ						tjɑ		
ʈ							ʈo												
d	di	dy	ɖɯ				do	da	dɑ	ɖə˞	dɣ					djɤ			
ɖ							ɖo												
nd		ndy	ɳɖɯ	(ɳɖu)[1]	(nde)[2]		ndo	nda	ndɑ	ɳɖə˞	ndɣ				ɳɖwə˞				
ɳɖ							ɳɖo												
l		ly	ɭɯ	ɭu	le		lo	la	lɑ	ɭə˞	lɣ								
ɭ							ɭo												
n	(ni)		ɳɯ		ne		no	na	nɑ	nə˞	nɣ					ɲɤ [3]			
ɳ							ɳo												

[1] Variante de /lu/.

[2] Variante de /le/.

[3] En diachronie, la combinaison [ɲɤ] peut provenir d'un ancien */njɤ/ aussi bien que d'un ancien */ŋjɤ/ ; au plan synchronique, le contraste entre /n/ et /ŋ/ est neutralisé dans cette position.

	i	y	ɯ	u	e	ɤ	o	a	ɑ	ə	ɣ	wɑ	wa	wɤ	wə	jɤ	jɑ	ja	jə
tsʰ		tsʰy	tsʰɿ (tsʰʐ)		tsʰe		tsʰo	tsʰa	tsʰɑ	tsʰə									
ʈʂʰ			ʈʂʰɻ = ʈʂʰʐ̩	ʈʂʰu		ʈʂʰɤ	(ʈʂʰo)			ʈʂʰə	ʈʂʰɣ	ʈʂʰwɑ			ʈʂʰwə				
ts	(tsi)	tsy	tsɿ (tsʐ)		tse		tso	tsa	tsɑ	tsə									
ʈʂ			ʈʂɻ = ʈʂʐ̩	ʈʂu		ʈʂɤ				ʈʂə	ʈʂɣ	ʈʂwɑ			ʈʂwə				
dz		dzy	dzɿ (dzʐ)		dze			dza											
ɖʐ̩			ɖʐ̩ɻ = ɖʐʐ̩	ɖʐ̩u		ɖʐ̩ɤ					ɖʐ̩ɣ								
ndz		ndzy	ndzɿ = ndzʐ̩				ndzo	ndza	ndzɑ										
ɳɖʐ̩			ɳɖʐ̩ɻ= ɳɖʐʐ̩							ɳɖʐ̩ə	ɳɖʐ̩ɣ	ɳɖʐ̩wɑ			ɳɖʐ̩wə				

	i	y	ɯ	u	e	ɤ	o	a	ɑ	əˑ	ɤ̣	wɑ	wa	wɤ	wəˑ	jɤ	jɑ	ja	jəˑ
(f)					(fe)[4]			(fa)[5]											
s	si	sy	sɻ = sʐ̩		se		so	sa	sɑ	səˑ						(sjɤ)			
ʂ			ʂɻ = ʂʐ̩	ʂu		ʂɤ	ʂo	ʂa		ʂəˑ	ʂɤ̣	ʂwɑ			ʂwəˑ				
z	zi	zy	zɻ = zʐ̩		ze		zo	za	zɑ										
ʐ			ʐɻ = ʐʐ̩	ʐu						ʐəˑ	ʐɤ̣	ʐwɑ			ʐwəˑ				
ç		çy																	
h	çi	hy	xɯ	xu	he	xɤ	hø	ha	xɑ	həˑ	fɤ̣	xwɑ				ç jɤ			
kʰ	tçʰi	tçʰy	kʰɯ	kʰu		kʰɤ	kʰo	kʰa	kʰɑ		kʰɤ̣	kʰwɑ						tçʰa	tçʰəˑ
k	ci	cy	kɯ	ku		kɤ	ko	ka	kɑ		kɤ̣	kwɑ				tçɤ		tça	tçəˑ
g	ɟi	ɟy	gɯ	gu		gɤ					gɤ̣					ɟjɤ		ɟja	
ŋg	ɲɟi	ɲɟy	ŋgɯ	ŋgu		ŋgɤ	ŋgo	ŋga	ŋgɑ		ŋgɤ̣			ŋgwɤ		ɲɟɤ		ɲɟa	ɲɟəˑ
ŋ	ɲi					ŋɤ			ŋɑ		ŋɤ̣					ɲɤ [6]			

Tableau 1. Réalisation phonétique des combinaisons initiales-rimes attestées dans le dialecte de AS. Les combinaisons entre parenthèses sont marginales. (Les questions posées par le découpage en initiale et rime, et la possibilité d'une réanalyse reconnaissant une médiale /w/, sont abordées au cours de l'analyse.)

[4] Emprunt au chinois.

[5] Emprunt au chinois.

[6] En diachronie, la combinaison [ɲɤ] peut provenir d'un ancien */njɤ/ aussi bien que d'un ancien */ŋjɤ/ ; au plan synchronique, le contraste entre /n/ et /ŋ/ est neutralisé dans cette position.

1. Les initiales

	bilabiales	coronales	rétroflexes	palatale	vélaires	glottale
occlusives	p^h p b m͡b	t^h t d n͡d	$ʈ^h$ ʈ ɖ ɳ͡ɖ		k^h k g ŋ͡g	
affriquées		t͡s^h t͡s d͡z n͡dz	ʈ͡ʂ^h ʈ͡ʂ ɖ͡ʐ ɳ͡ɖ͡ʐ			
nasales	m	n			ŋ	
fricatives		s z	ʂ ʐ	ç		h
approximantes latérale		l	ɭ			

Tableau 2. Phonèmes consonantiques du naxi occidental, dialecte de AS (qui appartient à l'ensemble des parlers occidentaux *du naxi)*

Le tableau 2 présente les phonèmes consonantiques du dialecte naxi de AS, adoptant les conventions de l'Alphabet Phonétique International (version de 1996).

Cet inventaire consonantique nous paraît être représentatif des dialectes occidentaux du naxi. Il est relativement plus étendu que celui d'autres parlers : les rétroflexes /ʈ^h ʈ ɖ ɳ͡ɖ ɭ/ sont absentes des dialectes de DYZ, NL et FK ; la distinction entre prénasalisées et voisées est perdue dans de nombreux dialectes dont DYZ et NL.

En revanche, la fricative palatale /ç/ paraît être récente (son origine est discutée ci-dessous, en section 1.1) ; elle est absente du dialecte de FK, plus conservateur sur ce point. Notre première notation de cette fricative était /ɕ/, suivant en cela la notation en usage dans les publications chinoises ; cette notation a été modifiée en /ç/ sur le conseil de Martine Mazaudon, qui nous a indiqué que /ɕ/ lui paraissait une notation inadéquate au plan phonétique dans le cas du phonème concerné.

Le tableau 3 présente les phonèmes consonantiques du dialecte de FK.

	bilabiales	coronales	rétroflexes	vélaires	glottale
occlusives	p^h p b m͡b	t^h t d n͡d		k^h k g ŋ͡g	
affriquées		t͡s^h t͡s d͡z n͡dz	ʈ͡ʂ^h ʈ͡ʂ ɖ͡ʐ ɳ͡ɖ͡ʐ		
nasales	m	n		ŋ	
fricatives		s z	ʂ ʐ		h
approximantes latérale		l			

Tableau 3. Phonèmes consonantiques du naxi occidental de FK (à la frontière avec le naxi oriental, ou mosuo*).*

Aspirées et prénasalisées ont été portées sur la ligne « occlusives », plaçant côte à côte quatre séries : aspirées, non aspirées, voisées et prénasalisées.

1.1. L'initiale palatale /ç/, fruit d'anciennes initiales nasales ? Oppositions de nasalité sur les voyelles et oppositions de consonnes fricatives

1.1.1. Description et mise en correspondance entre dialectes

Le dialecte naxi de FK présente trois rimes nasales : les voyelles antérieures fermées [ĩ] et [ỹ], qui apparaissent uniquement après une initiale fricative elle-même nasalisée, ci-après notée [ç̃] ; et la rime [ṽ] (en naxi, la fricative voisée /v/ est une rime, et ne peut jouer le rôle d'initiale), qui apparaît uniquement après [m̥] (son nasal labiodental non voisé). Sept exemples sont fournis dans le tableau 4.1 ; le décompte des items lexicaux qui possèdent la même composition phonémique que ces sept exemples n'a pas encore été établi.

Exemples de mots				
n°	FK	AS	traduction	traduction en chinois
1	ç̃ỹ ˋ	hyˋ	rouge	红
2	çjuˋ	çyˋ	fatigué	累
3	ç̃ĩ ˉ	çi ˉ	homme	人
4	çi ˉ	çi ˉ	riz (paddy)	稻子
5	m̥ṽ ˊ	fṽ	poil	毛
6	fṿ̀ té	*(emprunt au chinois)* fè	tombeau	坟墓
7	fṿ̀	fṽ	scie (probablement emprunt à la langue bai)	锯

Tableau 4.1 Transcription phonétique de syllabes à voyelles nasales du dialecte de FK et des mots correspondants dans le dialecte de AS. (Pour éviter la confusion entre diacritiques des tons et diacritique de nasalité, les tons sont décalés à droite des phonèmes.)

La correspondance phonétique est reprise dans le tableau 4.2.

dialecte de FK	ç̃ỹ	çju	ç̃ĩ	çi	m̥ṽ	fṿ
dialecte de AS	hy	çy	çi	çi	fṿ	fṿ

Tableau 4.2. Correspondance phonétique entre deux dialectes naxi pour cinq combinaisons initiale-rime.

La séquence [ju] de FK peut se réanalyser en /y/, et les initiales [ç] et [ç̃] en /h/ :

- à FK, /y/ se réalise [ju] en position initiale de mot, après fricative /h/, vélaire /k/ (qu'il palatalise) ; [y] après les coronales /nd d t t^h/, /l/ et /s/, et les bilabiales /mb b p p^h/. (Ainsi, « léger » se dit [jỳ] à AS et [jù] à FK ; l'une et l'autre forme peuvent s'analyser comme un simple /ỳ/.)

- à FK, [ç] est en distribution complémentaire avec [ɕ̃], [ŋ̊], [h], [f] et [x] : tous peuvent être réanalysés comme allophones de /h/ : /h/ est nasalisé dans les syllabes à voyelle nasale ; il adopte le lieu d'articulation de la rime : palatalisé devant les voyelles fermées d'avant, vélarisé devant les voyelles d'arrière fermé, et labiodental devant /ụ/. (Ce /h/ est distinct de l'« initiale zéro » ; au sujet des réalisations des syllabes sans consonne initiale, voir la première ligne du tableau 1.)
- Dans le dialecte de AS, /ç/ et /h/ ne s'opposent que devant /y/. (L'opposition est perdue chez le plus jeune de nos consultants, âgé de 28 ans à la date de la première enquête, qui prononce uniformément [çy].)

Le tableau 5 résume cette situation.

n°	FK	AS	traduction	traduction en chinois
1	hỹ ˋ	hyˋ	rouge	红
2	hyˋ	çyˋ	fatigué	累
3	hĩ ˉ	hi ˉ	homme	人
4	hi ˉ	hi ˉ	riz (paddy)	稻子
5	hụ̃ ˊ	fụ̄	poil	毛
7	fụ̀	fụ̄	scie	锯

Tableau 5. Une analyse phonémique possible des syllabes à voyelles nasales du dialecte de FK et des mots correspondants dans le dialecte de AS

Au plan phonémique, 1 et 2 se distingueraient par la rime à FK, par l'initiale à AS ; 3 et 4 se distingueraient par la rime à FK, tandis qu'ils sont confondus à AS (de même que 5 et 7)[1].

Cette situation synchronique complexe paraît appeler une réflexion diachronique : « A complex nasal state may best be understood in terms of the nasal process which gives rise to it » (Hyman 1975:254). **L'évolution diachronique de l'opposition peut être imaginée comme suit** :

(1) à un premier stade (hypothétique), /hỹ /, /hĩ /, /hỹ / s'opposaient à /hy/, /hi/, /hụ/ ; les réalisations phonétiques ne s'écartaient pas sensiblement de l'identité des phonèmes ([hỹ], [hy]) ;

[1] Une analyse concurrente consisterait à juger que les syllabes [ɕ̃ ỹ], [ɕĩ ˜] et [ŋ̊ṽ] de FK sont, au plan phonémique, /hy/, /hi/ et /hụ/, s'opposant à /ɕy/, /ɕi/ et /fụ/ ; cette dernière combinaison peut éventuellement être analysée, à un plan élevé d'abstraction, comme appartenant à la même série que les deux autres, et noté /ɕụ/. Cette analyse est artificielle, car il ne paraît alors pas possible de justifier l'apparition de la nasalité. La réanalyse de toutes les consonnes palatales comme des consonnes vélaires palatalisées (menée ci-dessous, section 2) conduit également à la conclure qu'il n'existait pas d'opposition /ɕ/ ~ /h/ à date ancienne.

(2) à un deuxième stade, l'opposition se marque avec régularité sur l'initiale, en plus de se marquer sur la rime : par exemple, dans l'opposition de /hỹ / à /hy/, réalisés, l'un comme [hỹ], l'autre comme [çy], la hiérarchie entre trait principal et trait secondaire n'est plus nette. C'est ce stade qui est attesté à FK, où /hĩ /, réalisé [hĩ], s'oppose à /hi/ (réalisé [çi]), et /hỹ /, réalisé [hỹ], à /hy/, réalisé (suite à une diphtongaison du /y/) comme [çju].

(3) à AS, l'opposition entre /hỹ / et /hy/ se conserve sous la forme d'une opposition d'initiales, devenant /hy/ ~ /çy/. L'opposition entre /hĩ / et /hi/ se perd sans compensation ; il n'existe à l'heure actuelle pas de moyen de déterminer si cette perte est ou non précédée d'une transphonologisation de l'opposition en une opposition entre initiales (/hi/ ~ /çi/).

Le tableau 6 résume ces hypothèses.

	Valeur phonétique des syllabes			
	Rimes en / y /		Rimes en / i / et / ɥ /	
Etape	dialecte de AS	dialecte de FK	dialecte de AS	dialecte de FK
état antérieur abstrait (reconstruit)	hỹ ~ hy		hĩ ~ hi, hɥ̃ ~ hɥ	
changement commun	hỹ ~ çy			hĩ ~ çi, hɥ̃ ~ fɥ
changements divergents	perte de nasalité > hy ~ çy	diphtongaison de /y/ > hỹ ~ çju	confusion en [çi] et [fv], respectivement	sans changement : hĩ ~ çi, hɥ̃ ~ fɥ

Tableau 6. Hypothèses sur l'évolution des syllabes à voyelle antérieure fermée après fricative glottale en naxi occidental.

Au plan de la réflexion sur les mécanismes du changement, **il paraît envisageable d'invoquer la nasalisation de la voyelle pour expliquer l'absence de palatalisation de la consonne initiale /h/ lorsqu'elle est suivie d'une voyelle nasale** : au stade 1 (hypothétique), l'abaissement du voile du palais avait sans doute lieu avant le début de la voyelle, par anticipation ; or l'ouverture du port vélo-pharyngien est défavorable à la réalisation d'une fricative [ç], car les fricatives nécessitent une pression d'air relativement élevée derrière la constriction, donc, dans le cas de [ç], une pression intra-orale relativement élevée. Du fait de l'ouverture du port vélo-pharyngien, la pression intra-orale ne peut guère s'élever : en cas de forte constriction orale (configuration nécessaire pour [ç]), l'impédance du conduit oral est modifiée, et l'air tend à s'échapper par le nez.

En revanche, les raisons pour lesquelles l'opposition se perd dans le cas de /i/ et non dans celui de /y/ n'ont pas été mises en lumière. Il ne nous paraît pas exister de motivation phonétique simple expliquant le fait.

L'évolution qui a eu lieu dans le dialecte de AS a pour effet, en synchronie, de donner statut phonémique au son [ç], cela dans un unique contexte : devant la voyelle /y/. Une application

à la lettre des principes de description synchronique amène à parler de neutralisation (de même qu'en français /e/ et /ɛ/ ne s'opposent qu'en syllabe ouverte, l'opposition étant neutralisée en syllabe fermée) : l'opposition entre /h/ et /ç/ est *neutralisée* dans tous les contextes, sauf devant /y/. (Au sujet de la notion de neutralisation, voir Trubetzkoy 1939 [1976]:80-87, Martinet 1969:257-259.) Au plan diachronique, le terme de « neutralisation » est contre-intuitif, puisque le statut distinctif acquis par /ç/ devant /y/ et dans ce seul contexte représente une innovation par rapport à un stade où [ç] était simple allophone de /h/ ; au plan synchronique, cette description paraît néanmoins adéquate, si extrême soit la « neutralisation » en question. Il paraît néanmoins avisé de présenter ces faits synchroniques dans leur contexte diachronique, pour ne pas induire en erreur les typologues[1].

Cette situation est similaire à celle des consonnes rétroflexes [ɳɖ ɖ ʈ ʈʰ], [ɳ] et [ɭ] (décrites ci-dessous, section 1.2), qui s'opposent aux alvéolaires /nd d t tʰ/, /n/, et /l/ devant /o/, opposition neutralisée devant les autres voyelles.

Notre choix est de noter /h/ dans tous les cas où l'opposition est neutralisée (par exemple : /hi/ pour [çi]), et de réserver la notation /ç/ aux contextes dans lesquels cette unité s'oppose à /h/, par exemple pour noter l'opposition entre /çỳ/ « fatigué » et /hỳ/ « rouge ».

Pour compléter ces observations, signalons une correspondance FK /hy/ ~ AS /ʂu/ (tableau 7). Celle-ci paraît être due à une confusion entre /ʂu/ et /hy/ (cette dernière syllabe étant réalisée phonétiquement [çju]) dans le dialecte de FK.

[1] Larry Hyman fournit un exemple spectaculaire de la nécessité de compléter la description synchronique par une réflexion diachronique : en wukari (Shimizu 1971) apparaissent des syllabes du type [ba], [bã], [mã], [m͡ba], à l'exclusion de [ma] et [mbã]. (Dans cette description, [b] représente une occlusive orale p, t, k ou b, d, g, [m] une nasale m, n ou ŋ.) Au plan synchronique, il est imaginable de réanalyser [mba] comme la réalisation d'un simple /ma/, [mã] étant simplement décrit, au plan phonémique, comme /mã/. L'économie réalisée en termes d'unités phonémiques postulées a pour prix des *règles* qui s'inscrivent en faux par rapport à des régularités typologiques bien établies, compliquant singulièrement le travail du typologue. Un examen diachronique révèle en réalité que m͡b est en fait un groupe anciennement attesté ; dans un état antérieur *(proto-jukunoïde)*, les structures syllabiques possibles étaient au nombre de neuf : *mam, *mab, *ma, *m͡bam, *m͡bab, *m͡ba, *bab, *bam, *ba. Quatre phénomènes ont ensuite eu lieu : les prénasalisées sont devenues des nasales géminées lorsque la syllabe comportait une finale nasale ; ces géminées se sont simplifiées en consonnes nasales ; les voyelles se sont nasalisées lorsqu'elles étaient précédées ou suivies d'une consonne nasale ; et enfin, les consonnes finales sont tombées (faits résumés par Hyman 1975:262).

FK	AS	traduction	traduction en chinois
/hỳ/ (phonétiquement [ɕju])	/ʂù/	fer ; *homophone* propre	铁
/hý kʰwù/ (phonétiquement [ɕju])	/ʂú kʰwù/	laid	丑
hý hȳ	ʂù, *réduplqué en* ʂū ʂù	chercher	寻找

Tableau 7. Illustration de la correspondance systématique entre syllabes /hy/ du dialecte de FK et syllabes /ʂu/ du dialecte de AS.

Le passage de [ʂu] à [ɕy] ou inversement est également attesté dans le domaine des dialectes chinois. Pour ne prendre qu'un exemple, le mot « tubercule » 薯 (*pinyin* shu[3]) est /ʂu[5]/ en chinois standard, /ɕy[5]/ dans le parler gan (贛) de Nanchang (nan[2] chang[1] 南昌) (il s'agirait d'un emprunt au mandarin de Hankou (han[4] kou[3] 汉口), plus conservateur sur ce point : voir Sagart 1993:121). Signalons également, à titre anecdotique, que certains Chinois anglicisent en [ʃuː] le nom de famille 徐, prononcé [ɕy] en mandarin : ainsi de Xu Yi, professeur à Londres, et auteur de travaux sur les tons mandarins.

1.1.2. Premières hypothèses diachroniques sur l'origine des voyelles nasales

L'analyse qui précède conduit à s'interroger sur l'origine des voyelles nasales du dialecte de FK. Le cas qui paraît le plus courant dans les langues est que la nasalité vocalique provienne d'anciennes consonnes nasales finales : la nasalisation d'une voyelle par une consonne nasale finale qui disparaît ensuite est un processus largement attesté, par exemple dans le passage du latin au français, et dans certains dialectes chinois qui ont perdu les nasales finales, par exemple celui de Qingdao (qing[1] dao[3] 青岛) ; au plan phonétique, l'influence d'une consonne nasale finale sur la voyelle précédente est bien documentée (voir par exemple Cohn 1990).

Cette première hypothèse ne trouve pas d'écho dans la comparaison avec d'autres langues tibéto-birmanes. Un rapide relevé montre que « rouge », /hỹ ˋ/ en naxi de FK, est /nɯ/ ou /ni/ dans les langues yi et en birman (birman écrit, et birman contemporain ; tons non notés) ; le fait que « rouge » se dise hong[2] (红) en chinois mandarin standard, avec une nasale finale, est de l'ordre de la coïncidence car le mot a longtemps signifié « rose » et non « rouge » (remarque due à William Baxter).

« Poil », /hṽ ˊ/ en naxi de FK, est /ɑ.mwe/ en birman, /m̥/, /mɯ/ ou /nu/ dans les langues yi. « Homme », /hĩ ˉ/ en naxi de FK, est /mi/ en tibétain écrit ; dans de nombreuses langues tibéto-birmanes, il s'agit pareillement d'une syllabe à initiale nasale et à voyelle /i/ (par exemple le /ni/ qui apparaît dans les ethnonymes des peuples *hani* et *sani*). **La présence de nasales initiales dans ces mots suggère l'influence d'une ancienne initiale nasale**.

Une consonne nasale initiale peut se dénasaliser de diverses façons. Elle peut évoluer vers une occlusive prénasalisée puis perdre sa prénasalisation (m > m͡b > b, n > n͡d > d, ŋ > ŋ͡g > g), comme c'est le cas dans certains dialectes chinois (Chen 1975). Cette évolution est en fait l'une des manifestations du phénomène plus général de la tendance à l'extension de la nasalisation au-delà de la consonne initiale : dans un parler qui oppose, par exemple, /ma/ à /mã/, l'évolution du /ma/ vers /mba/ peut représenter un moyen de bloquer la propagation de la nasalité du /m/ vers la voyelle, propagation qui menacerait la bonne conservation de l'opposition entre /ma/ et /mã/ (Larry Hyman parle de « perceptual reinforcement of the orality of a neighboring vowel » : Hyman 1975:256, 259). Le modèle pour comprendre l'évolution des faits naxi se trouve dans d'autres langues. La nasalisation d'une voyelle par une consonne nasale qui la précède est bien attestée dans les langues yi (aussi appelées ni, lolo) : les voyelles fermées (typiquement /i/ ou /u/) se trouvent fréquemment nasalisées, voire disparaissent tout-à-fait, après une consonne nasale (Bradley 1989, « Nasality as a prosody in Loloish »). En arakanais (dialecte birman), la voyelle ne disparaît pas mais se nasalise (Bradley 1985, « Arakanese vowels »).

Des données de langue nosu et phunoi (groupe ni/yi) amènent David Bradley à reconstruire un passage d'une consonne nasale dévoisée / ŋ̊ / à /h/, par la perte de l'occlusion orale, puis celle de la nasalité (Bradley 1979, *Proto-Loloish*) ; cette dernière étape s'explique aisément par le fait que l'effet acoustique de l'abaissement du voile du palais devient très faible en l'absence d'occlusion orale.

En tamang, une variation libre s'observe en synchronie entre initiale /ŋ/ suivie d'une voyelle orale et initiale /h/ suivie d'une voyelle nasale : « appeler » est tantôt réalisé [^{4}ŋot-pa], tantôt [^{4}hõt-pa] (le chiffre en exposant indique le ton lexical)[1].

Au plan phonétique, ces phénomènes sont notamment discutés par Ohala 1975 ; voir également Matisoff 1975, « Rhinoglottophilia: the mysterious connection between nasality and glottality »).

La lénition d'une nasale initiale (ou d'une nasale en position finale de groupe consonantique), son passage à /h/ (probablement par l'intermédiaire d'une nasale dévoisée) et la nasalisation concomitante de la voyelle suivante peuvent donc raisonnablement être postulés en naxi.

1.1.3. Données complémentaires pour la suite de la réflexion

Dans le dialecte de AS, l'initiale /ŋ/ ne se combine qu'avec trois rimes : /ɤ/, /ɑ/ et /ɣ/ ; l'idée selon laquelle les syllabes à initiale /ŋ/ étaient plus nombreuses dans un état ancien de la

[1] Communication personnelle de Martine Mazaudon.

langue reçoit un élément de confirmation de données de FK, où /ŋ/ se combine avec /u/ et /wɑ/, comme le montre le tableau 8.

FK	AS	traduction	traduction en chinois
ŋwɑ (*dans :* sá ŋwà⁻, ŋwà⁻ mē, ndā ŋwā)	wɑ (*dans :* sá wā, wá mē, ndā wā)	3e mois, 5e mois, 12e mois	三月，五月，十二月
ŋwà⁻	wā	tuile [sans doute emprunt au chinois]	瓦
ŋù	wà	verbe copule	是(这是什么)
ŋù⁻	kō	arroser	浇(浇菜)
ŋū	ŋɤ̄ (dans : ŋɤ̄ gɤ̄)	1e sg., possessif	我(的)
ŋà	ŋɤ̀	1e sg.	我
ŋɣ̣	ŋɣ̣	argent (le métal) ; *homophone* pleurer	银； 哭

Tableau 8. Correspondances entre dialectes de FK et AS pour les syllabes à initiale nasale vélaire.

Il est raisonnable de penser que la forme conservatrice est celle qui présente une nasale initiale. Ces faits rappellent l'évolution du wo[3] 我 du chinois mandarin standard contemporain : il proviendrait d'une forme */ŋajʔ/, postulée pour toute la période 1300-250 av. J.-C., de la période Shang tardive à la période des Zhou orientaux ; wu[2] 吾 proviendrait d'un ancien */ŋa/ (Sagart 1999:142-146 ; voir également Sagart 1993:123-124).

Une question qui se pose ensuite est de savoir pourquoi seules les voyelles nasales /ĩ /, /ỹ / et /ɤ̃ / existent actuellement. Le fait que les voyelles fermées se singularisent n'est pas propre au naxi : en môn également, elles présentent des correspondances inattendues (Christian Bauer, communication personnelle) ; en ce qui concerne plus particulièrement la nasalisation, Hajek et Maeda 2000 montrent que les voyelles basses n'ont pas nécessairement une plus grande propension à la nasalisation (complétant les observations de Chen 1975, Matisoff 1975:272) :

> ...there appear to be two competing universal tendencies, if not universals : one is of preferential nasalization of high vowels, and the other is of preferential nasalization of vowels with greatest duration (generally accepted to be low vowels. (p. 68)

L'examen de cette question nécessitera un examen systématique des cognats en naxi de syllabes à initiale nasale d'autres langues yi-birmanes.

1.2. Allophones rétroflexes des consonnes coronales, et rétroflexion phonémique

1.2.1. Description

L'inventaire des consonnes rétroflexes se limite, dans les dialectes naxi de NL et FK, à l'opposition entre /s/ et /z/ d'une part, /ʂ/ et /ʐ/ d'autre part. Les consonnes rétroflexes [ɳɖ ɖ ʈ ʈʰ], [ɳ] et [ɭ] sont en distribution complémentaire avec les alvéolaires /nd d t tʰ/,

/n/, et /l/ : les premières apparaissent devant les rimes /ɚ/, /ɯ/, et /u/, les secondes devant toutes les autres voyelles. (Le / ɻ / est proche du [ɽ] de l'API par la force de son relâchement.) Par exemple, « fermer (une porte) » se dit [ʈɚ́] ; il n'existe pas de combinaison phonétique [tɚ́].

En revanche, **dans le dialecte de AS, occlusives, nasales et latérales alvéolaires et rétroflexes s'opposent devant /o/.** Le tableau 9 présente des oppositions lexicales vérifiées auprès de quatre informateurs. Les exemples pour / ʈ / et /ʈʰ / ne sont pas des mots de base, mais ceux pour /nd - ɳɖ /, /n - ɳ /, /l - ɻ / paraissent l'être. La réalisation de la voyelle est quelque peu modifiée par la consonne initiale : le [o] est plus fermé et arrondi après les consonnes rétroflexes qu'après les consonnes alvéolaires.

initiale	ton	consonne non rétroflexe + /o/ 不卷舌辅音 + /o/	consonne rétroflexe + /o/ 卷舌辅音 + /o/
t ~ ʈ	L	/tò/ « obéir » ; /hĩ mɤ̃ tò/ : homme-NEG.-obéir « ne pas être obéissant » (不听话)	/ʈò/ (onomatopée) bruit d'une branche qu'on casse ; du maïs qu'on récolte. (采玉米或哲断树枝的声音)
	M	/tō/ « planche » (木板) ; /tō lō/ « chapeau de paille » (斗笠)	/ʈō/ « ampoule (sur les mains, les pieds) » ; /k^{h}i ʈō/ « bosse »
	H	/tó/ « front » (前额)	/ʈó lō/ « rouler » (卷起来)
t^{h} ~ ʈ^{h}	L	/t^{h}ò/ « être adossé à » (靠)	/ʈ^{h}ò/ « dormir comme une souche » (很深地睡着)
	M	/t^{h}ō/ « sapin » (松树) ; « lapin » (兔子)	/ʈ^{h}ō/ « frapper doucement » (敲) ; ex. : donner une tape à un enfant
	H	/t^{h}ó/ « incruster, sertir » (镶进去 : 牙.....)	/ʈ^{h}ó/ (onomatopée) bruit d'un petit objet qui tombe (petite pierre, par exemple) ; le ton varie selon l'informateur : ton moyen ou ton haut
n ~ ɳ	L	/nò/ « objet de commerce » (商品) : typiquement, les produits de la ferme	/ɳò/ « mite » (蛀虫) ; « brin (de fil) » (量词) 根 (一根线)
	M	/nō/ « herbes » (草，杂草)	/ɳō/ « lait » (奶) ; « sentir » (闻) ; « nom de famille » (姓)
	H	/nó/ (emprunt) « glutineux (ex. : riz glutineux ») 糯 (糯米) ; prendre, saisir	/ɳó ɳō/ « emmêlé (fil) » (线) 乱
nd ~ ɳɖ	L	/ndò/ « stupide » (愚蠢)	/ɳɖò/ « bobine » (量词) 团 (一团线) ; /ɳɖō ɳɖò çȳ çỳ/ « à la hâte, précipitamment » (急急忙忙)
	M	/ndō/ « grimper, gravir » (爬)	/ɳɖō/ « piquer, mordre (une abeille pique qqn) » (蜜蜂) 咬 ; « frapper à une porte, toquer à une porte »
	H	/ndó/ « tomber, faire une chute » (跌倒) ; /ō ndó/ « battre au fléau » ; /ndó ndzū/ « gober au vol, saisir au vol » (/ndzū/ : manger 吃)	/ɳɖó lō/ « un sac, une cargaison » (一小包，一驼) ; /ɳɖó/ « faire une boulette (de plastique, de papier, de fils...), rouler en boule » ; (onomatopée) bruit d'un petit objet qui vous heurte, par ex. une noix qui vous tombe sur la tête
l ~ ɭ	L	/lò/ « dans » (里面) ; « muntjac, *Muntiacus reevesi* (petit cerf d'Asie) » (麂子) ; « vallée » (山谷) ; « prendre en charge » (担得起)	/ū ɭò/ « joug » (轭) (/ū/ : vache)
	M	/lō/ « travail » (/lō bē/ « travailler » 工作 ; /bē/ « faire » 作) ; /tō lō/ « chapeau de paille » (斗笠) ; /ə lō/ « grand'père, aïeul »	/ʂō ɭō/ « os » (骨头) ; /ɭō/ « bouchée » (量词) 口 (一口饭) ; /tʂ^{h}ó ɭō/ « fourmi » (蚂蚁) ; /hi ´ɭō/ « muet, personne muette »
	H	/ló/ « enjamber, franchir » (跨 : 跨小沟) ; et emploi grammaticalisé : /ló lē bū/ « s'en retourner » (回去) (/lē/ « à nouveau », /bū/ « aller »)	/ɭó pi ` / « piment » (辣椒)
d	L-M-H	do ~ ɖo : pas d'oppositions ; combinaison attestée : /do/, ex. : /dò/ « voir » (见).	

Tableau 9. Exemples de mots lexicaux illustrant les oppositions entre initiales rétroflexes et non rétroflexes devant /o/ dans le dialecte de AS.

Signalons que le /o/ n'a pas d'allophones après les occlusives coronales qui empiète sur le domaine phonétique des voyelles voisines : il n'y a pas confusion entre les syllabes du

tableau 9 et telle ou telle des syllabes constituées d'une voyelle /ɯ/, /u/, /ə/ précédée d'une occlusive coronale (dont la réalisation est rétroflexe devant ces voyelles). Pour prendre un exemple (notant l'archiphonème par une majuscule) : /Lū/, impératif du verbe « venir » (réalisé [[lū]), ne se confond pas avec /l̩ō/, « joug » (réalisé [l̩ō]).

1.2.2. Analyse synchronique

En synchronie, se pose la question de l'attribution de l'opposition à l'initiale (analyse présentée dans les tableaux ci-dessus) ou à la rime.

Un professeur d'université issu du village de AS, He Xueguang (和学光教授), a pris conscience des distinctions lexicales portées par ces oppositions syllabiques ; il a ajouté une rime *-uo* distincte de *-o* dans la romanisation naxi (dont la version standard est présentée par He Jiren et Jiang Zhuyi 1985 ; He Zhiwu 1987), qu'il enseigne dans le cadre du Centre de diffusion de la culture naxi qu'il a fondé à Lijiang. Le tableau 10 illustre les équivalences entre une transcription d'orientation phonétique, la notation de He Xueguang, et l'analyse phonémique proposée ici.

approximation en API	notation orthographique proposée par He Xueguang	analyse phonémique
l̩ɔ ~ lo	luo ~ lo	l̩o ~ lo
ɳ̩ɔ ~ no	nuo ~ no	ɳ̩o ~ no
etc.		

Tableau 10. Proposition de notation des oppositions entre syllabes avec et sans rétroflexion initiale par He Xueguang.

L'analyse en termes de différences sur la rime présente un avantage d'économie : elle permet de n'ajouter à l'inventaire phonémique qu'une unité.

Néanmoins, au plan de l'économie du système dans son ensemble, la reconnaissance d'une quatrième voyelle d'arrière arrondie (choix de Fu Maoji 1981) amènerait à ajouter un quatrième degré d'aperture vocalique dans un système à trois degrés, introduisant en outre une asymétrie entre trois degrés pour les voyelles d'avant et quatre pour celles d'arrière, ce qui serait une configuration inhabituelle au plan phonétique.

Cet argument non plus n'est pas décisif car les tendances majoritaires n'ont pas force de loi (comme le rappelle sur un plan général Nolan 1999, « The devil is in the detail », et comme l'illustrent des systèmes vocaliques tels que celui de la langue bora, étudié par Parker 2001, qui oppose, parmi les voyelles d'arrière, /ɨ/, /ɯ/ et /o/).

Un dernier argument est apporté par la comparaison avec les langues apparentées : en birman, la présence de voyelles mi-fermées et mi-ouvertes (/e/ et /ɛ/, /o/ et /ɔ/) est une innovation du birman central ; les dialectes arakanais et marma ne distinguent que trois

degrés d'aperture vocalique (Denise Bernot, communication personnelle). Cet argument non plus n'est pas décisif, car deux dialectes proches peuvent avoir des systèmes vocaliques nettement différents (par exemple, les différences entre français méridional et français parisien, ou entre « received pronunciation » et accents du nord de l'Angleterre, sont plus fortes au plan vocalique qu'au plan consonantique).

Au plan phonétique, l'interprétation comme une rime /wo/ nous paraît exclue, et l'interprétation comme une rime /ɔ/ problématique ; la différence vocalique nous paraît beaucoup moins nette que la différence entre initiales. **Ces réflexions vont dans le sens d'une description phonémique en termes d'*initiale différente*.**

Au plan synchronique, comme dans le cas de l'opposition entre /h/ et /ç/, qui n'est présente que devant /y/, et est neutralisée dans tous les autres contextes, l'opposition entre coronales rétroflexes et non rétroflexes représente un cas extrême de neutralisation. Elle est neutralisée dans tous les contextes sauf un : avant la voyelle /o/.

1.2.3. Comparaison entre dialectes

Les oppositions entre occlusives coronales rétroflexes et non rétroflexes n'existent pas dans les parlers de la ville de Lijiang et de sa périphérie immédiate : « lait » et « herbe » y sont homophones (/nō/), de même que « grimper » et « mordre/piquer » (/ndō/), « dans » et « joug » (/lò/). Dans certains cas, la voyelle des mots correspondant aux deux formes qui s'opposent à AS est néanmoins différente : ainsi, « piquer, mordre (une abeille pique quelqu'un) », /ɳɖō/ à AS, est /ndə̄/ chez un informateur de la ville de Lijiang, /dɤ̄/ chez une informatrice plus jeune (30 ans environ) qui habite également Lijiang, dont sont originaires ses deux parents (la coexistence et le mélange de dialectes à Lijiang n'a rien de surprenant, s'agissant de la principale ville de la région naxi).

Dans le dialecte de NL, une même voyelle apparaît dans tous les mots correspondant aux mots du dialecte de AS qui comprennent un /o/ après coronale non rétroflexe (colonne de gauche du tableau 9) : il s'agit de la voyelle /o/ (prononcée [ø]). Des voyelles diverses correspondent aux /o/ après rétroflexes : principalement /ɑ/, mais également /o/ et /u/. Le tableau 11 présente quatre correspondances.

initiale	AS	NL	correspondance	sens du mot
ɳ	ɳò	nà	o ↔ ɑ	mite 蛀虫
	ɳō	nō	o ↔ o	lait 奶
ɳɖ	ɳɖò	ndu˩	o ↔ u	CL. des fils, des bobines (量词) 团 (一团线)
	ɳɖō	ndā	o ↔ ɑ	piquer, mordre (une abeille pique) (蜜蜂)咬
ɭ	ɭó pi˴	lā pi˴		piment 辣椒
	ʂō ɭō	ʂā lā		os 骨头
	ɯ̄ ɭò	ɯ̄ là		joug 轭

Tableau 11. Illustration des correspondances entre les syllabes à initiales rétroflexes de AS et les cognats dans le dialecte de NL

L'opposition entre occlusives coronales rétroflexes et non rétroflexes est perdue à NL, les syllabes en question devenant homophones d'autres syllabes : ainsi, « joug », /là/, y est homophone de « main » (qui se dit /là/, à NL comme à AS). Seule la mise en correspondance des dialectes permet de déceler la trace d'une ancienne opposition.

Li Lincan, Zhang Kun *et al.* 1953 (étudiant le parler de He Cai 和才, originaire de lu[3] dian[4] 鲁甸) notent également une différence entre l'initiale /l/ et une initiale /r/ qui correspond à notre / ɭ / ; néanmoins, ces auteurs demeurent au stade d'une notation phonétique : ainsi, ils notent [rə́˞] « graine » (/lə́˞/), malgré l'absence d'opposition avec [lə́˞], tandis qu'ils notent [lā] « tigre », pareillement en l'absence d'opposition. De même que nos / l / et / ɭ /, leur /l/ et /r/ ne s'opposent que devant /o/[1].

Les données rapportées par Fu Maoji 1981:297-307 révèlent le même type d'opposition phonémique. Ses données sont fondées sur le parler de 和芳 He[2]Fang[1], né dans le village de Zhonghe (zhong[1] he[2] 中和村) (Fu Maoji 1981:8) ; aux dires des linguistes He Jiren 和即仁 et He Limin 和力民, cet informateur aurait pratiqué plusieurs dialectes, et aurait notamment été influencé par le parler de Yangxi (yang[4] xi[1] 漾西). Une enquête auprès de He Jiren, lui-même natif de Yangxi, montre la présence de cette opposition dans son parler.

La comparaison entre dialectes naxi ne permet pas d'hypothèses sur l'origine de cette série de rétroflexes. A l'avenir, une comparaison systématique du naxi avec le birman, les langues ni/yi et les langues qiang du sud pourra fournir des indices.

[1] Dans le détail, les données ne se recoupent que partiellement. Les syllabes qu'ils notent [lo] correspondent toutes à « notre » /lo/ ; en revanche, certaines syllabes qu'ils notent [ro] correspondent à /lu/ dans nos données, par exemple [rō] « quatre », /lū/ dans nos données (prononcé [ɭū]). Nous n'avons pu vérifier leur mot pour « Sichuan » : [ró pʰā dỳ], désigné par un mot d'emprunt chez nos informateurs.

1.3. Allophones palatalisés des vélaires devant /i/, /y/, /jɤ/, /ja/, /jɑ/, /jə/

1.3.1. Description générale du phénomène

Plusieurs descriptions phonémiques du naxi (dont He Jiren et Jiang Zhuyi 1985, Fu Maoji 1981:297-316, Fang Guoyu et He Zhiwu 1995) suivent en partie les notations en usage dans les descriptions linguistiques du chinois mandarin, employant les notations /tɕʰ/ /tɕ/ /dʑ/ /ndʑ/ pour les consonnes palatales[1]. Cette notation présente l'inconvénient de donner une fausse idée du découpage en initiale et rime : /dʑɤ̀/ « courir » et /gɤ̀/ « aubergine » sont analysés comme ayant pour rime /ɤ/, alors que celle-ci est en fait un /jɤ/ dans le premier cas : /gjɤ̀/, la semi-voyelle /j/ étant précisément cause de la palatalisation de l'initiale vélaire. Réécrite comme /gjɤ̀/ ~ /gɤ̀/, **l'opposition phonémique s'avère être une opposition entre rimes et non entre initiales**.

La notation de ces auteurs a en outre pour effet de masquer certaines oppositions. Les consonnes occlusives vélaires /kʰ/, /k/, /g/, /ŋg/ sont systématiquement palatalisées devant les rimes /i/, /y/, /jɤ/, /ja/, /jɑ/ et /jə/. Pour prendre l'exemple de la rime /jɤ/, les combinaisons /tʰjɤ/, /tjɤ/, /djɤ/, /ndjɤ/ s'opposent à /kʰjɤ/, /kjɤ/, /gjɤ/, /ŋgjɤ/. La notation qui accorde statut phonémique aux affriquées palatales ne permet pas cette distinction : ainsi, /tɕɤ/ est employé par ces auteurs pour deux combinaisons en réalité distinctes, /tjɤ/ et /kjɤ/. (Des exemples sont fournis dans le tableau 12 : par exemple, /kjɤ́/ « faire bouillir », /tjɤ́/ « plier (des vêtements...) ».) L'initiale /ʑ/ notée par ces auteurs est un /j/ dans certains cas (le mot noté /ʑɑ̄ kò/, phonétiquement [jɑ̄ kò], est /jɑ̄ kò/), la réalisation phonétique d'une attaque syllabique douce dans les autres cas (les syllabes notées /ʑi/, phonétiquement [ji], sont au plan phonémique de simples /i/ : voir ci-dessus la première ligne du tableau 1).

Cette réanalyse, qui rejoint la notation de Martine Mazaudon et Boyd Michailovsky (Mazaudon et Michailovsky 1979), permet de faire l'économie des initiales /tɕʰ/, /tɕ/, /dʑ/, /ndʑ/ et /ʑ/. Cette analyse est convaincante en diachronie, et offre un point de départ fiable dans l'entreprise de reconstruction d'états anciens de la langue. En synchronie, la situation est moins évidente : au risque de nous hasarder à des prédictions sans utilité scientifique, il nous semble que l'évolution future de la langue pourrait prêter une réalité phonémique à l'analyse qui calque la notation officielle du chinois mandarin standard[2].

---

[1] Les notations tɕʰ et tɕ sont les équivalents des *q* et *j* de la romanisation *pinyin*. En revanche dʑ et ndʑ ne correspondent à aucun phonème chinois, mais sont employés en suivant l'analogie avec tɕʰ et tɕ, le point d'articulation des sons naxi transcrits par ces quatre transcriptions étant le même.

[2] Une forme expressive indique que les initiales palatalisées sont d'un certain point de vue des sons distincts des vélaires dont elles sont allophones : il s'agit d'une variante ludique de /ʈʂʰù/ « vite, rapide » ; elle est notée phonétiquement entre crochets dans l'énoncé ci-dessous :

1.3.2. Rôle de la langue nationale dans la palatalisation

La palatalisation sous l'effet des voyelles d'avant fermées est un phénomène très répandu (voir notamment les articles « La palatalisation du roman septentrional » et « La palatalisation 'spontanée' de g en arabe » dans Martinet 1975, *Evolution des langues et reconstruction*) ; il n'est pas nécessaire d'invoquer des facteurs externes pour expliquer son apparition. Il paraît en revanche clair que la pression de la langue nationale tend à accélérer cette évolution. Dès l'école primaire, les jeunes Naxi sont rompus à un découpage des syllabes en initiale, rime et ton selon le système de translittération *pinyin* (pin^{1}yin^{1} 拼音), qui note séparément les alvéolo-palatales : *j* pour [tɕ], *q* pour [tɕ^{h}], distincts de [k] (noté *g*) et [k^{h}] (noté *k*) ; les jeunes locuteurs qui ont à cœur de maîtriser le mandarin paraissent avoir tendance à utiliser les mêmes sons lorsqu'ils parlent naxi, ce qui rejoint les observations d'André Martinet : les systèmes phonologiques restent distincts, tandis qu'une modification phonétique peut avoir lieu dans les sons qui se ressemblent dans les deux langues en contact.

> Le bilinguisme hautevillois ne semble pas avoir deux systèmes articulatoires complètement différents pour le français et pour le patois, comme c'est souvent le cas chez ceux qui parlent, depuis leur tendre enfance, deux langues de civilisation comme le français et l'anglais. Ceux-ci n'articuleront pas de façon identique le fr. *chou* et l'anglais *shoe*. Celui-là, au contraire, ne fera aucune différence entre le fr. coup et son équivalent patois ku, dans ce sens que la prononciation du mot français sera adaptée aux habitudes locales, ce qui entraîne un [k] particulièrement ferme et un [u] un peu antérieur. Ceci a naturellement comme corollaire que si l'Hautevillois modifie le détail de son articulation du français, que ce soit par suite de séjours faits hors de Savoie, ou par imitation des « Parisiens » qu'il peut entendre, l'articulation du patois suivra la même voie. (...) le parallélisme que nous signalions peut, dans certains cas individuels, recevoir quelques entorses: certains sujets pourront parfois, en français, faire un effort vers le « beau-parler », tandis qu'ils conserveront en patois des articulations plus naturelles. Il reste néanmoins qu'en général les nuances phoniques tendent à être identiques dans les deux idiomes, tandis que les deux systèmes phonologiques demeurent bien distincts. (Martinet 1956:53)

Quelques exemples illustreront l'état actuel de la palatalisation. Le linguiste He Jiren 和即仁, locuteur natif qui vit depuis plus de cinquante ans dans la capitale de la province

---

(1)	dɯ̄ mə̀	[tɕ^{h}i˗ tɕ^{h}ō]	bē	lū !
	un peu	terme expressif	particule adverbialisante	venir (impératif)

« Dépêche-toi de venir ! »

L'analyse phonémique proposée ci-dessus aboutit à /k^{h}i˗ k^{h}jō/, forme asymétrique ; en revanche, si (au rebours de notre analyse phonémique) la forme expressive est analysée en /tɕ^{h}/ + /i/ et /tɕ^{h}/ + /o/, le jeu revient à une alternance entre une voyelle simple d'avant et une voyelle simple d'arrière, ce qui s'observe dans des langues très diverses. Le fait que l'initiale de ces syllabes ne soit pas /ʈʂ^{h}/ comme dans le cas du mot de base, /ʈʂ^{h}ù/, tient à ce que les fricatives et affriquées rétroflexes ne se combinent pas avec les voyelles d'avant /i/, /y/, /e/ en naxi (voir l'inventaire des combinaisons attestées dans le tableau 1 ci-dessus).

(Kunming), dans un environnement chinois, paraît réaliser de façon totalement identique, au plan phonétique, les palatalisées naxi et chinoises. Les jeunes locuteurs (par exemple F1 et M1) semblent suivre le même chemin. Au plan dialectal, le parler de Lijiang (丽江) paraît très avancé dans la direction d'une identification aux sons chinois, tandis que les dialectes de AS et FK conservent une articulation des consonnes en question qui est plus en arrière qu'en mandarin. Le dialecte de Longquan (long^{2}quan^{2} 龙泉) paraît conservateur de ce point de vue, d'après la description de Thomas Pinson (Pinson 1996 ; nous n'avons pas encore eu l'occasion d'étudier ce dialecte).

L'écriture naxi romanisée, qui propose le même découpage que le *pinyin* (pin^{1}yin^{1} 拼音) chinois, n'est pas enseignée ni employée en dehors de cercles très restreints ; son influence sur la langue parlée peut être considérée à l'heure actuelle comme nulle.

1.3.3. La palatalisation a-t-elle abouti à la perte d'oppositions syllabiques ?

Les oppositions syllabiques sont globalement maintenues au cours du processus de palatalisation ; en revanche, **une confusion a pu avoir lieu entre d'anciens */ŋjɤ/ et */njɤ/**, à supposer que ces deux syllabes existaient préalablement à la palatalisation : les deux se seraient alors confondues pour donner les actuelles syllabes [ɲɤ]. **Au plan synchronique, la neutralisation de l'opposition entre initiales nasales /n/ et /ŋ/ conduit à adopter la notation phonémique /Njɤ/**, dans laquelle /N/ représente un archiphonème nasal qui s'oppose à /m/ (exemples : /mjɤ̀/ « œil », /kā Njɤ̀/ « jadis »).

Il est imaginable que l'actuelle combinaison [ɲi] corresponde, non seulement à d'anciens */ŋi/, mais également à d'anciens */ni/ qui auraient également subi la palatalisation. Au plan synchronique, la rareté des syllabes [ni] (voir détails ci-dessous) fournit un argument dans ce sens : suite à la palatalisation des anciens */ni/, la case vide laissée dans le système aurait été remplie par quelques formes qui demeurent marginales. En lisu, où un ancien */ni/ a évolué vers une réalisation phonétique [ɲi], c'est la combinaison /ne/ qui a rempli la case phonétique laissée vide, et se réalise [ni] (Bradley 1979, 1994).

En naxi de AS, la syllabe /ni/ apparaît marginalement, dans des formes dont certaines sont incontestablement des formes expressives : une onomatopée /ni ˊni ˊnə̀ nə̄/ (mentionnée par Fang Guoyu et He Zhiwu 1995, non attestée dans nos données) ; des noms propres et mots enfantins, tels que /ə̄ ni ˋˊ/ « petit cochon », lesquels ont un statut à part[1] ; mais aussi certains mots lexicaux : /kɤ̄ ni ˋ/ « cédrèle de Chine » (Fang Guoyu et He Zhiwu 1995:151) ; /ni ˊnə̀/ « instrument à vent semblable au hautbois » (suo^{3} na^{4} 唢呐) ; /k^{h}ɣ̣ ni ˋˊ/ « se cacher à l'intérieur de quelque chose ». Une comparaison entre dialectes et entre langues sera

[1] Sur le phénomène de remplissage de case par les formes expressives, voir par exemple Haudricourt 1952.

nécessaire pour apporter des arguments en faveur de notre hypothèse selon laquelle l'actuelle catégorie que nous analysons comme /ŋi/ (réalisée [ɲi]) a été en fait alimentée, au plan diachronique, par la palatalisation d'anciens */ni/, les actuels /ni/ étant le fruit d'évolutions ultérieures.

2.2. Données détaillées concernant les rimes /jɤ/ et /jo/

Les paragraphes qui suivent sont spécifiquement consacrés aux rimes /jɤ/ et /jo/, les oppositions d'initiales devant ces rimes n'étant pas transcrites dans les principaux travaux antérieurs.

La rime /jo/ est très marginale en naxi, et n'est donc pas portée dans le tableau 1. Elle apparaît exclusivement sur des mots empruntés au chinois, avec une seule exception : un mot qui semble appartenir aux franges expressives du lexique[1]. Les mots « neuf (9) » (九) et « vin » (酒) sont homophones en mandarin standard contemporain (jiu[3]), mais sont respectivement reconstruits, au stade du moyen-chinois, comme /kjuw/ et /tsjuw/ (Baxter 1992, *A handbook of Old Chinese phonology*), distinction qui ne s'est perdue qu'à date récente en mandarin standard, et demeure dans plusieurs dialectes du mandarin (guan[1] hua[4] 官话) et dans certains styles oratoires (ainsi dans l'opéra de Pékin, où la distinction est appelée fen[1] jian[1] tuan[2] 分尖团). La réalisation de ces mots comme /kjō/ et /tjō/ dans la variété de chinois apprise par nos informateurs naxi indique que ces mots demeurent distincts dans le dialecte chinois du Sichuan avec lequel nos informateurs sont en contact.

Le tableau 12 présente un certain nombre de mots illustrant les oppositions phonémiques en question.

[1] Mots chinois mis à part, un seul mot possède la rime /jo/ : « fourmi », noté /tɕʰō lò/ par He Jiren et Jiang Zhuyi 1985, peut se réanalyser en /kʰjō lò/ (prononciation de nos informateurs de la ville de Lijiang) ; ce mot paraît posséder des cognats plausibles dans des langues tibéto-birmanes (tibétain *grog-ma*, japhug /qro/, somang /kʰorók/, zbu /qʰrôχ/ ; données de Jacques 2004) ; néanmoins, un petit animal côtoyé au quotidien est particulièrement sujet à singularité phonétique, et de fait le mot est prononcé /ʈʂʰó ɭō/ à AS, /ʈʂʰā ŋ̍ɣ̍/ à FK, ce qui ne témoigne pas d'une présence stable et régulière de la rime /jo/ en naxi.

	rime /jɤ/		rime /jo/	
	initiale vélaire palatalisée	initiale coronale	initiale vélaire palatalisée	initiale coronale
occlusive pré-nasalisée	/ŋgjɤ̃/ « pénible, dur » (困难)	/ndjɤ̃/ « joli, beau » (漂亮) ; /kū mū ndjɤ́ ndjɤ̃ t^{h}ā/ « rajuster son chapeau »	(non attesté)	(non attesté)
occlusive voisée	/gjɤ̃/ « courir » (跑)		(non attesté)	(non attesté)
occlusive non voisée	/kjɤ́/ « faire bouillir » (煮) ; /kjɤ̄/ « lèpre » (麻风) ; /kjɤ̃/ « être en mesure de » (担得起) (/lō bē mɤ̃ kjɤ̃/ : travail-faire-NEG.-être en mesure de : ne pas être en mesure de faire un travail)	/kā tjɤ́ tjɤ̄/ « plier des vêtements, ranger des vêtements » (叠) ; /tjɤ̃ tjɤ̃/ : soucis, contrariétés (麻烦)	/kjō/ : prononciation du mot chinois « neuf » (九)	/tjō/ : prononciation du mot chinois « vin, alcool » (酒) ; /tjó/ : prononciation du mot chinois « sauver » (jiu^{4} 救) ; ex. : /ŋɤ̃ kỳ tjó tjō lū/ « protégez-moi ! »
occlusive aspirée	/k^{h}jɤ́/ « afficher, coller (une affiche) » 贴(贴标语)	/t^{h}jɤ́/ « goutte » (滴) ; /t^{h}jɤ̄/ « couler, dégouliner » (流) ; /t^{h}jɤ̃/ « vallée » (山谷)	/hi^{-} k^{h}jǒ/ mot chinois « merle » xi^{3}que^{4} (喜雀)	(non attesté)

Tableau 12. Exemples de mots lexicaux illustrant les oppositions entre syllabes à initiales vélaires palatalisées et syllabes à initiale coronale de même mode d'articulation.

Au plan phonétique, **les initiales coronales devant /jɤ/ et /jo/ (de même que devant /i/) sont réalisées apicales et sans affrication, tandis que les vélaires palatalisées (laminales) sont fortement affriquées.** L'initiale de /kjo/ est moins palatalisée que celle de /kjɤ/ ; l'arrondissement de la rime /jo/ est moins net que celui de la rime /o/.

Il peut être utile de signaler que les rimes /jɤ/ et /jo/ ne sont en distribution complémentaire avec aucune des autres rimes. Ainsi, la rime /ɯ/ se combine librement avec les initiales coronales : /ndù/ « abattre (un arbre) », /dū/ « un (1) », /tū/ « se lever » (dans : /gɤ̃ tū/), /t^{h}ū/ « 3^{e} pers. sg. » ; sa réalisation phonétique correspond directement à son identité phonémique ([ndù], [dū], [tū], [t^{h}ū] ; de même pour les autres consonnes coronales).

Une confirmation de l'analyse phonémique du mot /ndjɤ̃/ « joli » est apportée par la comparaison entre dialectes : il se prononce /ljɤ̃/ dans le dialecte de DYZ. Dans le dialecte de AS (principal dialecte utilisé ici), l'initiale /nd/ correspond pour partie à des /nd/ de DYZ, pour partie à des /l/ de DYZ : ainsi, « terre » (地) est /ndū/ à AS, /lū/ à DYZ. Le découpage des syllabes /ljɤ̃/ (à AS) et /ndjɤ̃/ (à DYZ) est donc clairement le suivant : /l/ + /jɤ/, /nd/ + /jɤ/, avec une rime /jɤ/.

Cette même analyse apparaît en fait de façon sporadique dans les travaux publiés sur le naxi, sous forme de notations qui sont erronées au regard des notations préconisées par la

transcription officielle. Ainsi, la notation /tɕɤ/ est préconisée (notamment par He Zhiwu 1987) pour notre syllabe /kjɤ/ ; dans un petit livre de proverbes (纳西谚语-科空，云南民族出版社，郭大烈/郑卫东主编, 1998, p. 14), dont les notations sont faites par un locuteur naxi possédant un certain entraînement à l'Alphabet Phonétique International, le mot « grenouille » (que nous analysons comme /pā kjɤ̄/) est noté [pā tɕiə̄], ce qui revient à une analyse en : initiale [tɕ] + rime [iə], laquelle coïncide précisément avec la rime que nous notons /jɤ/. De même, bien que Pinson 1998 note les palatalisées comme une série distincte (par exemple [tɕi] pour /gi/), certaines de ses notations révèlent la tentation d'une réanalyse : l'intensif « très », que nous analysons comme /gjà/ (suivant en cela Mazaudon et Michailovsky 1979), est noté [dʑjà] par Pinson. Dans ses transcriptions du dialecte de Longquan (long^{2}quan^2 龙泉), visiblement moins avancé sur le chemin de la palatalisation, Pinson note les consonnes concernées comme des vélaires : ainsi pour /kjá/, classificateur des véhicules (initiale vélaire, rime /jɑ/), que He Jiren et Jiang Zhuyi 1985 notent /tɕá/. Enfin, une notation (isolée) du dictionnaire de Li Lincan, Zhang Kun et He Cai (Li Lincan, Zhang Kun *et al.* 1953) revient pareillement à adopter une analyse en initiale vélaire + rime commençant par une semi-voyelle /j/ : « se chauffer au soleil » est noté /m͡bù.ŋ͡gié/, sans la palatalisation, alors que la cohérence avec les autres notations de l'ouvrage demanderait /m͡bù.ndʑé/ ou /m͡bù.ndʑó/ (ce qui est la notation adoptée par Fang Guoyu et He Zhiwu 1995:102).

1.3.4. Choix de conserver un statut distinct à certains /ɯ/ modifiés

Certains morphèmes comportant la rime /jɤ/ paraissent être dérivées de morphèmes qui présentent une voyelle fermée /i/ ou /ɯ/.

Ainsi, /gjɤ̄ kʏ̀/ « sur le toit (de la maison) » (房顶上) et /hī dū gjɤ́/ « une famille » (一家人) illustrent la combinaison /g/ + /jɤ/ ; il paraît clair qu'ils sont dérivés de /gi`/ « maison », par des processus qui ne sont plus productifs (peut-être l'ajout d'un affixe). De même, dans /ts^{h}è djɤ̄/, qui désigne le 11^e mois de l'année (十一月), on reconnaît /ts^{h}è/ « dix » et /dū/ « un » (note : « onze » se dit /ts^{h}ē dù/, avec changement tonal). Enfin, /djɤ̄/ dans le superlatif /zē djɤ̄ djɤ̄/ « très grand » est à rapporter à /dù/ « grand ». Du point de vue strictement synchronique, ces mots auraient leur place dans le tableau 12 ; néanmoins, le caractère catégoriel du changement n'a pu être établi avec une pleine certitude, c'est pourquoi ces formes n'ont pas été intégrées au tableau 12, bien qu'il soit imaginable que leur évolution future les fasse entrer pleinement dans le jeu d'opposition phonémique qu'illustre ce tableau.

En revanche, dans d'autres noms de mois, des morphèmes qui sont visiblement dérivés de chiffres se distinguent de la forme de base par une **diphtongaison centralisante finale dont le résultat phonétique ne coïncide avec aucune rime qui ait statut phonémique dans la langue**. Les noms de mois sont pour la plupart construits, comme en chinois, selon le schéma

nombre + mot « mois » (par exemple : /ŋi˧/ « deux », /mē/ « mois », /ŋi˧ mē/ « le deuxième mois, février »). Le onzième, mentionné ci-dessus, n'est pas seul à connaître une modification de sa voyelle ; deux autres comportent une diphtongaison centralisante finale sur le chiffre, comme le montre le tableau 13.

mois	nombre correspondant	forme composée
4e mois	**lū**	**lúᵊ mē**
9e mois	**ŋgɤ̄**	**ŋgūᵊ mē**
11e mois	tsʰè « dix » + dù « un »	tsʰè djɤ̄

Tableau 13. Modification des nombres dans les noms de mois (dialecte de AS)

Ce changement a pour effet de distinguer « quatre mois » et « le quatrième mois », « neuf mois » et « le neuvième mois ». La même différenciation a lieu dans le dialecte de FK pour le nom du 4e mois, [luᵊ ˋ ˉ mē] (par souci de clarté typographique, les diacritiques qui transcrivent le ton lexical LH de la syllabe [luᵊ] sont notés à la suite de celle-ci, et non en les y superposant).

La rime des syllabes [lúᵊ] et [ŋgūᵊ] ne peut être identifiée à /wɤ/ (ce qui donnerait /lwɤ́ mē/, /ŋgwɤ̄ mē/). Les formes modifiées demeurent transparentes pour les locuteurs : elles se définissent par contraste avec la forme de base, non par une réalisation stable et réellement intégrée au système phonologique. La modification pourrait aboutir, soit à l'entrée des rimes en question dans une catégorie vocalique préexistante (le plus probable étant un passage à /wɤ/), soit à la création d'une rime nouvelle, telle que /uʷ/.

1.4. Autres observations concernant les initiales

1.4.1. Allophones « trills » des occlusives bilabiales /pʰ/, /p/, /b/, /m͡b/ devant /ɥ/

He Jiren et Jiang Zhuyi 1985:9 signalent que l'association de la rime /ɥ/ à une initiale bilabiale se réalise avec une vibration des lèvres (« trill »), que les auteurs notent par l'ajout d'un [ʳ]. Nos observations confirment que la réalisation phonétique de ces syllabes comporte une particularité : **l'intégralité de la syllabe est bilabiale ; en d'autres termes, la rime adopte le lieu d'articulation de l'initiale**. Le tableau 14 présente ces faits, adoptant la notation par un B majuscule en petites capitales, [ʙ], proposée par l'Alphabet Phonétique International (version de 1996) pour « voiced bilabial trill » (à ne pas confondre avec la fricative bilabiale [β]).

Phonèmes	Réalisation phonétique, d'après He Jiren et Jiang Zhuyi 1985	Nouvelle proposition de transcription	traduction en français	traduction en chinois
$p^h ɣ́$	$p^{hr} ɣ́$	$p^h ɓ́$	semer à la volée	撒(撒种子)
$pv́$	$p^r ɣ́$	$pɓ́$	offrir	送
$bɣ̄$	$b^r ɣ̄$	$ɓ̄$	casserole	锅
$\widehat{mb}ɣ̀$	$\widehat{mb}^r ɣ̀$	$mbɓ̀$	ramper	爬(虫在地上爬)

Tableau 14. Propositions de notation de la réalisation phonétique des syllabes à initiale occlusive bilabiale et à rime /ɣ/ dans le dialecte naxi de AS.

1.4.2. Allophones uvulaires des vélaires devant les voyelles d'arrière

Les voyelles occlusives vélaires sont réalisées plus en arrière devant les voyelles d'arrière, se rapprochant de [q^h q ɢ ɴɢ], illustration supplémentaire de la tendance à la coarticulation à l'intérieur d'une même syllabe.

1.4.3. Fricatives

La fricative glottale /h/ est fortement coarticulée avec la voyelle qui suit, comme le montrent les phénomènes décrits ci-dessus, section 1.1. Elle est réalisée avec une friction principalement glottale, [h], devant /ɚ/, /e/, /o/, /a/ ; elle est vélaire ([x]) devant les voyelles d'arrière /ɯ/, /u/, /ɤ/, /ɑ/ ; palatale [ç] devant /i/, /y/, et les rimes commençant par une semi-voyelle palatale ; labiodentale [f] devant /ɣ/. Devant les rimes /i/ et /y/, les fricatives /h/ et /ç/ s'opposent : ainsi, /hỳ/ « rouge » s'oppose à /çỳ/ « fatigué ». Comme il a été exposé en détail ci-dessus, section 1.1, cette opposition marginale s'explique par le transfert à l'initiale d'une opposition entre voyelles orales et voyelles nasales, conservée dans le dialecte de FK, où « rouge » se dit /hỹ`/, et « fatigué » se dit /hy `/ : **lors de la perte des oppositions de voyelles orales et nasales dans le dialecte de AS, la variation allophonique du /h/ initial, palatalisé devant les voyelles orales** (phénomène qui s'observe actuellement à FK)**, a acquis statut phonémique**. Notre choix est de noter /h/ dans tous les cas où il n'y a pas d'opposition (par exemple : /hɣ/, /hi/), et de réserver la notation /ç/ aux contextes dans lesquels cette unité s'oppose à /h/, par exemple dans l'opposition entre /çỳ/ « fatigué » et /hỳ/ « rouge ». Le son [ç] est donc *diphonémique*, à l'instar des rétroflexes [ɳɖ ɖ ʈ ʈʰ], [ɳ] et [ɭ], qui sont de simples allophones de /nd d t t^h/, /n/, et /l/ devant les voyelles /ɚ/, /ɯ/, /u/, /ɤ/ mais s'opposent aux alvéolaires non rétroflexes devant /o/ (voir section 1.2).

Le /h/ est noté /x/ par He Jiren et Jiang Zhuyi 1985. Le phonème figure ainsi dans une série où il voisine avec la nasale /ŋ/ et les occlusives /k^h/, /k/, /g/, /ŋg/, tandis que la notation /h/ le

fait apparaître seul membre d'une série glottique. Néanmoins, la notation /x/ est difficile à justifier phonétiquement : il est aisément compréhensible que /h/ ait un allophone vélaire devant une voyelle d'arrière comme /u/, tandis que l'apparition d'un allophone [h] d'un phonème /x/ devant une voyelle d'avant comme /e/ ne serait pas clairement motivée.

Si l'on accepte l'hypothèse selon laquelle [fy̩] est la réalisation phonétique de /hy̩/ (hypothèse qui a pour elle l'analogie avec les faits en lisu[1], mais ne saurait être considérée comme démontrée), **le phonème consonantique /f/ est probablement dû à des emprunts récents au chinois** (notamment /fā/ pour la syllabe très fréquente fa[1] 发 « envoyer, devenir »). (La langue bai, bai[1]yu[3] 白语, voisine du naxi et qui a peut-être exercé une influence sur les dialectes occidentaux du naxi, doit le phonème /f/ au chinois, selon Xu Lin et Zhao Yansun 1984, qui signalent qu'à la date de l'enquête /f/ était encore absent de deux des dialectes.)

A la différence du chinois mandarin, le naxi possède une opposition entre consonnes voisées et non voisées ; il a acquis un /v/, aux côtés du /f/, par lequel il interprète les /w/ initiaux de syllabes chinoises : ainsi, wen[3] 稳 « stable » est emprunté par le naxi comme /vē/ (avec disparition de la nasale finale, comme cela a systématiquement lieu lors d'emprunt de mots chinois, du fait de l'absence de consonnes finales en naxi) ; « moustiquaire », wen[2]zhang[4] 蚊帐, devient /vē tsá/. Ce /v/ n'existe pas tel quel en chinois ; il n'en est pas moins perçu comme un phonème propre à la langue dominante, puisqu'il apparaît par des mots empruntés au chinois ; ce statut s'accompagne d'un prestige qui paraît lui valoir une extension progressive dans le lexique, mot par mot (par un mécanisme décrit en section 3.2.4) : /ý̩ zi¯/, « oiseau », est prononcé [vi´ zi¯] par deux de nos consultants (F1, et M4, chez qui cette prononciation est en variation libre avec [y̩´ zi¯] et [vɯ´ zi¯]), prononciation qui nous paraît provenir de l'extension, mot par mot, de l'emploi du phonème perçu comme prestigieux. Une syllabe /və̃/ est également attestée, dans le sens « trempe, courage » : /t^{h}ɯ̄ və̃ gȳ/, ou /t^{h}ɯ̄ və̃ ī/ (3^{e} sg.-trempe, courage-prédicat d'existence) signifie « il a de la trempe, il a du cran ». L'absence de syllabe /wə̃/ dans le système suggère que [və̃] pourrait être la réalisation de cette combinaison ; mais aucune certitude n'a pu être atteinte quant à son origine.

1.4.4. Nasales

Aux dires du consultant M3, qui a habité dans plusieurs régions de langue naxi, certains dialectes de la plaine de Lijiang (par exemple autour de Qihe (qi[1] he[2] 七河) et Changshui

---

s[1] En lisu, /hu/ se réalise [fu], et il s'agit là de la seule combinaison dans laquelle le son [f] apparaisse dans cette langue, mots d'emprunt mis à part (Bradley 1994, *A Dictionary of the Northern Dialect of Lisu*).

(chang[2] shui[3] 长水, anciennement Qinglong qing[1] long[2] 青龙) connaîtraient une confusion de /l/ et /n/. Le phénomène n'est pas sans exemple : dans la variété populaire du vietnamien de Hanoi, /l/ et /n/ tendent actuellement à se confondre en /l/ ; dans la prononciation de Nam Định, près de Hanoi, /l/ et /n/ tendent à se confondre en /n/. Ainsi, la collation prise au lever, /lé nɣ́/, deviendrait /né nɣ́/ (exemple fourni par He Jiren et Jiang Zhuyi 1985:11, mais sans indication du dialecte concerné). Cette confusion n'a pas lieu dans les dialectes que nous avons étudiés.

1.4.5. Latérale l, et confusions avec n͡d

Dans le dialecte de AS, certaines syllabes qui possèdent une initiale /l/ dans le dialecte de DYZ (et d'autres dialectes occidentaux) ont tendance à connaître un changement de leur initiale en /n͡d/ devant /ɯ/, /u/ et /ɚ/ : ainsi, « terre », /lūˈ/ à DYZ, est fréquemment prononcé [n͡dūˈ] par le locuteur M4, parfois prononcé [lūˈ] ; « barque » (船) est plus souvent prononcé /n͡dùˋ/ que /lùˋ/, prononciation du dialecte de DYZ. Le chiffre « quatre », homophone de l'unité de mesure /lūˈ/ (两 : 一两银子), est prononcé, tantôt [n͡dūˈ], tantôt [lūˈ] ; « devoir » est tantôt /lɚ̄/, tantôt /ndɚ̄/. Le remplacement des /l/ par des /n͡d/ dans ces contextes n'est pas systématique : ainsi, aucune réalisation [n͡dūˈ] de /lūˈ/ « venir » n'a été observée.

1.4.6. Occlusives et affriquées : description en quatre degrés de *Voice Onset Time*, et recoupements possibles avec l'opposition *fortis/lenis*

Une possibilité ouverte pour la modélisation phonologique des consonnes occlusives et affriquées du naxi consiste à les décrire comme quatre séries distinguées par le moment où s'établit le voisement, rapporté à l'instant du relâchement de l'occlusion orale (en anglais, ce paramètre est désigné comme VOT : *Voice Onset Time*). Cette description reviendrait à pousser à ses limites la proposition de Lisker et Abramson (Lisker et Abramson 1964, « A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements » ; Abramson et Lisker 1970, « Discriminability along the voicing continuum: Cross-language tests ») qui consiste à ramener aspiration et voisement à la chronologie relative du début des vibrations glottiques et du décollement des organes en occlusion. Abramson et Lisker (suivis par Cho et Ladefoged 1999, « Variation and universals in VOT: evidence from 18 languages ») envisagent trois valeurs : aspiré, non voisé, et voisé. (Les idées de Lisker et Abramson sont discutées par Chomsky et Halle 1968:164, qui les rejettent, et par Householder 1971:55.) Les prénasalisées du naxi pourraient se plier à ce traitement : elles représenteraient un degré extrême de VOT négatif ; la réalisation par une prénasalisation, plutôt que par une occlusive voisée de tenue significativement plus longue que pour la série /b/, /d/, /g/, tiendrait à la difficulté de maintenir le voisement pendant une occlusion longue : le fait de garder le port vélo-pharyngien ouvert au début de la consonne serait une conséquence phonétique de la chronologie du début du voisement.

Ainsi, le voisement commence avant le relèvement du voile du palais pour le premier degré de VOT (ou « degré –2 », par rapport au temps de relâchement de l'occlusion orale), celui des prénasalisées /mb/, /nd/, /ŋg/ ; avant le relâchement de l'occlusion orale pour le deuxième degré (ou « degré –1 »), celui des voisées /b/, /d/, /g/ ; au relâchement de l'occlusion pour le troisième degré (« degré 0 »), celui des non voisées /p/, /t/, /k/ ; et après le relâchement de l'occlusion pour le quatrième degré (« degré +1 »), celui des aspirées /p^h/, /t^h/, /k^h/. Cette formulation aux allures de paradoxes est suggérée par l'évolution des distinctions lors des confusions entre prénasalisées et voisées qui ont eu lieu dans de nombreux dialectes naxi occidentaux (dont NL et DYZ) : ces deux séries se confondent en prénasalisées (parfois réalisées, à l'intérieur d'un énoncé, comme de simples voisées) ; **la direction prise par cette confusion revient à maximiser la distance en termes de VOT entre les trois séries qui demeurent**.

Comparaison entre les occlusives naxi et françaises : occlusives non voisées et occlusives fortis

S'agissant des voisées /b/, /d/, /g/ et des non voisées /p/, /t/, /k/, les études menées dans diverses langues sur le temps d'établissement du voisement tendent à montrer qu'au plan phonétique, les frontières entre « voisées » et « non voisées » ne coïncident pas d'une langue à l'autre (voir notamment les travaux de Lisker et Abramson cités ci-dessus). Lors des premières séances d'élicitation de vocabulaire (auprès des consultants F1, M1, M2, M3), je notais voisée la série non voisée : /bɯ/ au lieu de /pɯ/, /tsɑ/ au lieu de /dzɑ/, confondant en voisées toutes les occlusives et affriquées non aspirées et non prénasalisées, au point de douter de la conservation de cette opposition (rapportée par des travaux antérieurs) dans le parler de ces consultants. Cela paraît indiquer que les consonnes non voisées du français, ma langue maternelle, se distinguent des sourdes du naxi. Il est possible de décrire cette différence en termes de VOT : les non voisées du français se caractériseraient par un VOT positif (légère aspiration), tandis que les non voisées du naxi auraient un VOT nul. Klaus Kohler (communication personnelle) propose de décrire l'opposition des deux catégories de consonnes françaises, voisées et non voisées, comme une opposition entre *lenis* et *fortis*, catégories qui sont actuellement délaissées au profit du paramètre de VOT, lequel paraît plus objectif du fait que sa mesure est relativement simple, à la différence de l'exploration expérimentale des nuances articulatoires. (Sans entrer dans le détail de ce domaine passionnant que nous connaissons peu, signalons que les études articulatoires confirment qu'il existe des différences au plan supraglottique entre les séries de consonnes qui se distinguent par leur « VOT ».) Les occlusives non voisées du naxi n'étant pas *fortis*, à la différence des françaises, elles sonnent comme des voisées pour une oreille française.

Le paragraphe qui suit rapporte une observation qui confirme que le temps d'établissement du voisement n'est pas le seul corrélat qui distingue les quatre séries d'occlusives en naxi.

Perte fréquente du voisement pendant l'occlusion des occlusives et affriquées prénasalisées

Les enregistrements électroglottographiques réalisés simultanément avec l'enregistrement audio au long des séances d'enquête indiquent de façon très précise à quel moment le voisement commence et s'interrompt. Ces enregistrements montrent une **fréquente interruption du voisement après la partie nasale des initiales occlusives et affriquées prénasalisées, /mb/, /nd/, /ŋg/, /ndz/, /ɳɖʐ/**. Le fait pourrait se noter en alphabet phonétique international de la façon suivante : /ŋg/ a tendance à se réaliser [ŋk], /ɳɖʐ/ à se réaliser [ɳʧʂ]. L'interruption du voisement n'est pas systématiquement présente.

Au plan fonctionnel, l'interruption du voisement ne nuit pas à la reconnaissance de ces consonnes, que la présence de prénasalisation, complétée du lieu d'articulation, suffit à distinguer. Dans le cas des affriquées, dont la partie fricative est, de par les lois de l'aérodynamique, peu compatible avec un voisement optimal, le dévoisement de la partie fricative est particulièrement fréquent.

Ces phénomènes naxi rappellent des observations en langue bai (voisine du naxi) et en français. Dans le dialecte bai de Bijiang (bi^{4} jiang^{1} 碧江), les occlusives voisées ont évolué vers des prénasalisées qui ne sont pas voisées pendant l'occlusion : l'opposition entre /p/, /t/, /k/ d'une part, /b/, /d/, /g/ de l'autre est devenue chez certains locuteurs une opposition entre /p/, /t/, /k/ d'une part, /mp/, /nt/, /ŋg/ de l'autre (Xu Lin et Zhao Yansun 1984 ; l'enquête sur laquelle repose cette publication a eu lieu à la fin des années 1950). Dans les parlers nasu du centre-nord du Yunnan (localités de Luquan et Wuding), les occlusives prénasalisées voisées telles que /m͡b/ ont pareillement évolué vers des occlusives prénasalisées non voisées, voire aspirées : [m͡p] ou [m͡p^{h}] (David Bradley, communication personnelle).

En français, des groupes pré-nasalisés [m͡b], [n͡d], [ŋ͡g] apparaissent dans certains réalisations hyper-articulées de /b/, /d/, /g/ (J. Vaissière, cours de commentaires de spectrogrammes). Le degré de pré-nasalisation serait fonction de la difficulté de maintenir le voisement pendant la consonne occlusive : g/ŋ͡g sera le plus pré-nasalisé, d/n͡d sera assez fortement nasalisé, b/m͡b le moins pré-nasalisé. A la lumière des données naxi, il sera intéressant d'étudier l'éventuelle corrélation entre présence de prénasalisation et dévoisement pendant l'occlusion.

2. Observations synchroniques sur les rimes

Les rimes simples sont représentées sur la figure 1.1. Les rimes à semi-voyelle /jɑ/, /jɤ/, /ɥe/, /wa/, /wɑ/, /wɤ/, /wə/ ont été placées à proximité de leur voyelle principale (par

exemple, pour /wɑ/, la voyelle /ɑ/). La figure 1.2 présente des données expérimentales pour les voyelles simples, et comporte les allophones [z̩̍] et [z̩̀] de /ɯ/.

La voyelle /ɯ/ a pour allophone /z̩/ après les fricatives et affriquées coronales (/s/, /z/, /ts^{h}/, /ts/, /dz/, /ndz/), **et /z̩̍/ après les fricatives et affriquées rétroflexes** (/ʂ/, /ʐ/, /ʈʂ^{h}/, /ʈʂ/, /ɖʐ/, /ɳɖʐ/) : pour simplifier, la rime est le prolongement de la partie fricative de l'initiale. En chinois mandarin, les mêmes sons, /z̩̀/ et /z̩̍/ (pour lesquels Chao Yuen-ren propose des symboles spéciaux, [ɻ̩] et [ɹ̩], non reconnus par l'Alphabet Phonétique International), sont allophones de /i/, dans les mêmes contextes.

2.1. Syllabes labialisées : rime /wəˠ/, initiales labialisées, ou médiale /w/ ?

Les syllabes composées d'une initiale coronale (/nd/, /d/, /t/, /t^{h}/, /ɳɖʐ/, /ɖʐ/, /ʈʂ/, /ʈʂ^{h}/, /ʂ/, /ʐ/) suivie de la voyelle /əˠ/ dans les dialectes de DYZ et NL ont deux correspondances à AS, dont l'une comporte un arrondissement médial. Ainsi, DYZ/NL /ʂə̄ˠ/ « sept » est pareillement /ʂə̄ˠ/ à AS, tandis que « plein (un récipient est plein) », DYZ/NL /ʂə́ˠ/ (Mazaudon et Michailovsky 1979:31, He Jiren et Jiang Zhuyi 1985), est /ʂ^{w}ə́ˠ/ à AS. Le tableau 15 présente les oppositions relevées.

initiale	syllabe sans arrondissement	syllabe avec arrondissement
occlusive coronale aspirée	/tʰə˞/ : non attesté	/tʰʷə́˞ tʰʷə̄˞/ « emballer » (forme réduplicative de /tʰʷə́˞/)
		/tʰʷə̄˞/ : non attesté
		/tʰʷə̄˞ tʰʷə̀˞/ « mélanger » (搀) (forme réduplicative de /tʰʷə̀˞/)
occlusive coronale non voisée	/tə́˞/ « fermer (une porte, ...) »	/tʷə́˞/ « faire une vannerie » (/mú tʷə́˞/ « faire une vannerie en bambou »)
	/tə̄˞/ : non attesté	/tʷə̄˞/ : attacher (拴)
	/tə̀˞/ « fantôme » (魔鬼)	/tʷə̀˞/ « planche à découper » (菜板)
occlusive coronale voisée	/də˞/ : non attesté	/dʷə́˞/ : nom propre (nom de personne)
		/dʷə̄˞/ « étagère » (架子)
		/dʷə̀˞/ « bulle ; faire des bulles, bouillonner »
occlusive coronale prénasalisée	/ndə́˞/ : non attesté	/ndʷə́˞/ (onomatopée) son du gong, des cymbales
	/ndə̄˞/ : verbe de modalité « être nécessaire, falloir » ; interchangeable avec /lə̄˞/	/ndʷə̄˞/ « étang » ; « faux » (dans : /ndʷə̄˞ ŋi ´/)
	/ndə̀˞/ « appeler (à grands cris) »	/ndʷə̀˞/ : non attesté
affriquée rétroflexe aspirée	/ʈʂʰə̄˞/ « laver »	/ʈʂʰʷə́˞ « poumon » ; « saisir, agripper » /là ʈʂʰʷə́˞/ « arme » (枪) ; /ʈʂʰə̄˞/ : dans /zō ʈʂʰʷə̄˞ mē/ « belle-fille » (/zō/ « fils », /mē/ « mère, femme ») ; /ʈʂʰʷə́˞ ʈʂʰʷə̄˞/ « petit pain à la vapeur » (馒头)
affriquée rétroflexe non voisée	/ʈʂə́˞/ « tige » ; /là ʈʂə́˞/ « bras »	/ʈʂʷə̄˞/ « tousser » (咳嗽), aussi dans /ʈʂʷə̄˞ pé/ « glaire, crachat » (痰)
affriquée rétroflexe voisée	/ɖʐə˞/, /ɖʐʷə˞/ : non attestés	
affriquée rétroflexe prénasalisée	/ɳɖʐə˞/ : non attesté	/ɳɖʐʷə̄˞/ « nager » ; « humide », aussi dans /ɳɖʐʷə̄˞ kʰi ´/, /ɳɖʐʷə̄˞ pʰə̀˞/ « rosée » (露) (/kʰi ´/ « froid », /pʰə̀˞/ « blanc »)
fricative rétroflexe voisée	/ʐə̄˞/ « avoir peur » ; /ʐə̀˞/ : onomatopée	/ʐʷə̄˞ tʰē/ « couteau » ; variante : /zū tʰē/
fricative rétroflexe non voisée	/ʂə̄˞/ « sept ; chose » ; /ʂə́˞/ « fichu » (头帕)	/ʂʷə́˞/ « remplir à ras bord ; plein à ras bord » ; « petite collation »
approximante latérale	pas d'opposition ; syllabe attestée : /lə˞/, réalisée /ɭə˞/, et interchangeable avec /ndə˞/ dans de nombreux cas (la confusion des deux initiales est presque complète)	

Tableau 15. Exemples de mots lexicaux illustrant les oppositions entre syllabes à initiale coronale et voyelle /ɚ/ avec et sans arrondissement dans le dialecte de AS.

Dans sa version adaptée de la romanisation naxi, He Xueguang a ajouté une rime *-vr* pour noter ces syllabes. Sa prononciation est particulièrement rétroflexe et fermée ; elle pourrait se noter [ʷɻɪ], précisant que les deux articulations sont simultanées : la rime est légèrement arrondie de part en part.

Li Lincan, Zhang Kun *et al.* 1953 notent une finale [ur] distincte de la finale [ɚ] (notée [ʌr]). (Rappelons que le parler qu'ils étudient est celui de lu^{3} dian^{4} 鲁甸.) Cette rime comportant arrondissement est globalement en correspondance avec les /ʷɚ/ de AS. Faute de connaissance de première main de ce dialecte, il n'est pas encore possible de déterminer s'il s'agit phonétiquement d'un [uɚ], ou d'une réalisation proche de celle observée à AS.

Fu Maoji 1984:304 rapporte une syllabe [l̥uʌɹ], illustrée par les mots « se réunir, se rassembler » (聚集) et « en désordre, en bouleversement » (慌乱), mais sans fournir de mots prononcés [l̥ʌɹ] qui attesteraient de l'opposition d'arrondissement[1]. L'informateur de Fu Maoji, He Fang (和芳), vient du village de Zhonghe (zhong^{1} he^{2} 中和), mais, au dire du linguiste naxi He Jiren (communication personnelle), il a de la famille à Yangxi (yang^{4} xi^{1} 漾西) et aurait adopté certains traits de prononciation de ce village. Chez He Jiren, lui-même originaire de Yangxi, l'opposition d'arrondissement est présente, d'une façon qui correspond aux observations faites à AS ; ayant pris l'habitude de la notation standard proposée pour le naxi, qui ne note pas l'arrondissement, He Jiren n'indique pas cet arrondissement dans ses notations.

Comme dans le cas des oppositions étudiées plus haut, celles-ci posent la question de l'attribution de la différence à l'initiale ou à la rime, ou encore de la reconnaissance de l'existence de semi-voyelles médiales dans la structure syllabique. Les notations proposées par Fu Maoji 1984 et Li Lincan, Zhang Kun *et al.* 1953 reviennent à adopter une analyse en termes de rimes différentes. Les premières correspondances établies entre dialectes ne nous paraissent pas apporter d'arguments en faveur de l'une ou l'autre analyse. Les correspondances avec le dialecte de FK montrent, tantôt une rime /ɚ/ sans arrondissement (« rosée » est /ɳʤʐʷɚ̃ kʰi ˊ/ à AS, /ɳʤʐɚ̃/ à FK comme à DYZ), tantôt un changement de voyelle (/ɤ/, /wɑ/) et un ton haut. Le tableau ci-dessous présente trois de ces correspondances.

[1] D'après ce que nous savons des notations de cet auteur, le sommet de syllabe serait le /ʌɹ/, que nous interprétons, d'après la comparaison avec nos propres données, comme un /ɚ/ ; en effet, à nos notations /w/ et /j/ dans les noyaux vocaliques complexes correspondent respectivement /u/ et /i/ chez Fu Maoji.

initiale	AS	FK	correspondance	sens du mot
ʐ̩	ʐ̩ʷə̠ ndzà	ʐ̩ɤ́ ndzə̠	ʷə̠ ↔ ɤ́	saule, 柳树
nd	gi ˋ ndʷə̠	ndwɑ́	ʷə̠ ↔ wɑ́	étang, bassin 池塘
ɳɖʐ̩	ɳɖʐ̩ʷə̠	ɳɖʐ̩ə́	ʷə̠ ↔ ə́	humide, 湿
	ɳɖʐ̩ʷə̠	ɳɖʐ̩ə̠	ʷə̠ ↔ ə̠	rosée, 露

Illustration des correspondances entre les syllabes labialisées de AS et les cognats dans le dialecte de FK

Remarquons au passage que le dernier exemple, « rosée », fournit un contre-exemple à la correspondance entre une syllabe au ton moyen dans le dialecte de AS et une syllabe à ton haut dans le dialecte de FK, laquelle aurait éventuellement pu être interprétée comme signe de la présence d'anciens groupes de consonnes initiales.

La solution provisoirement retenue consiste à reconnaître l'existence de /w/ comme semi-voyelle médiale, jugeant que le naxi possède une structure syllabique (C)(G)V, où C représente une consonne, G (pour *Glide*) une semi-voyelle /w/ ou /j/, et V une voyelle, seule composant indispensable de la syllabe. La distribution des semi-voyelles médiales est très limitée : l'une et l'autre n'apparaissent que devant /ɑ/, /a/, /ɤ/ et /ə̠/, et après certaines consonnes seulement. **Cette analyse revient à diminuer de huit unités** (wɑ, wa, wɤ, wə̠, jɑ, ja, jɤ, jə̠) **l'inventaire des unités vocaliques**, au prix de la reconnaissance d'une structure syllabique plus complexe que la simple association d'une initiale et d'une unité vocalique. Ce choix permet également de réanalyser comme /wə̠/ la combinaison [və̠] (marginalement présente dans le lexique, voir section 1.4.3), réanalyse qui ne présente toutefois pas un degré complet de certitude.

2.2. La fricative /ɣ̩/, noyau de syllabe

En naxi, /ɣ̩/ est une rime, au même titre que les autres sons représentés sur la figure 1.2. Le fait que des consonnes voisées puissent avoir statut de noyau de syllabe est bien établi : ainsi de /n/ et /l/ en anglais (dans *button, bottle*, réalisés [bʌtn̩], [bɒtl̩]), /r/ et /l/ en sanskrit (dont les emplois vocaliques et les emplois consonantiques sont distingués dans la graphie indienne) et en tchèque (*prst* « doigt », *plný* « plein » ; exemples cités par Martinet 1986)[1].

Au plan phonétique, après les initiales non voisées, l'amplitude du signal électroglottographique augmente au fil de la rime [ɣ̩], tandis qu'il a tendance à décroître pour les autres rimes. Cette observation indique que le voisement se met en place plus

[1] Au sujet des consonnes sommet de syllabe, il peut être intéressant de noter, pour élargir les perspectives typologiques, qu'un mot entier peut ne comprendre aucun phonème voisé, comme le montre l'exemple du berbère de Tashlhiyt. Les premières descriptions des « mots sans voisement » de cette langue, solidement argumentées au plan phonologique, ont récemment reçu une confirmation expérimentale (voir Dell et Elmedlaoui 1985 « Syllabic Consonants and Syllabification in Imdlawn Tashlhiyt Berber », Dell et Elmedlaoui 2002 *Syllables in Tashlhiyt Berber and in Moroccan Arabic*, et Ridouane 2004, « Les mots sourds en berbère chleuh : analyses fibroscopiques et photoglottographiques »).

difficilement pour [ɣ], illustration du compromis aérodynamique à réaliser pour les fricatives voisées : en effet, au plan aérodynamique, les exigences de la production du voisement et de la production du bruit de friction sont contradictoires : pression intra-orale basse pour que le voisement soit possible (celui-ci demandant une différence de pression transglottique), pression intra-orale élevée pour que le bruit de friction soit produit (Jacqueline Vaissière, cours de phonétique fonctionnelle ; les premières observations expérimentales sur ce phénomène auraient été faites par Kenneth Stevens). A l'oreille, il semble que la friction s'affaiblisse fréquemment à mesure de la rime, au profit du voisement. Les notations de Mazaudon et Michailovsky 1979:24, notations phonétiques étroites, reflètent pareillement l'impression d'une **friction qui s'affaiblit parfois en fin de syllabe** : /lɣ̄ bɣ̣/ « petit-fils » est noté /lɣ̄ bvə̄/ ; /hē k^hý/, « boucle d'oreille », est noté /hē k^hvə́/.

He Jiren et Jiang Zhuyi 1985:9 notent également que la rime /ɣ̣/ se réalise souvent comme une approximante [ʋ], plutôt que comme une véritable fricative. Après les initiales nasales, le /ɣ̣/ est réalisé nasalisé ; la notation [ṽ] est en réalité équivalente, au plan phonétique, avec [ɱ] : les combinaisons /mɣ̣/, /nɣ̣/, /ŋɣ̣/ se réalisent [mɱ], [nɱ], [ŋɱ], le sommet de syllabe étant le [ɱ].

Proximités phonétiques de /ɣ/

Au plan phonétique, il paraît intéressant de déterminer de quelle voyelle le /ɣ̣/ (instable comme toute fricative voisée) se rapproche lorsque sa friction s'affaiblit. Cette réflexion doit contribuer à déterminer la relation du /ɣ̣/ aux autres rimes.

Lors du travail de transcription, nos erreurs les plus fréquentes consistaient à noter /u/, parfois /ɯ/. Certaines notations de Mazaudon et Michailovsky 1979 vont dans le même sens : /ī gū dỳ/ pour « la région naxi », pour notre /ī gɣ̄ dỳ/ ; /nú mē/ « cœur », /nɣ́ mē/ dans nos données. Changement catégoriel chez les consultants de Mazaudon et Michailovsky 1979, ou impression erronée de l'enquêteur ? Dans l'un et l'autre cas, le fait fournit un indice : **le /ɣ̣/ est proche des voyelles d'arrière**.

La variation dialectale paraît également révélatrice des affinités phonétiques du /ɣ̣/, même s'il est imprudent de relever des correspondances isolées avant d'avoir établi les correspondances entre dialectes pour l'ensemble des phonèmes, travail qui reste à faire. « Ver de terre » (蚯蚓) est /ə bū ʐɣ́ lū/ chez F1, M3 et M4, /ə bɣ̄ ʐɣ́ lū/ chez M2, alors qu'il n'existe pas de correspondance régulière entre le /ɣ̣/ du second et le /ɯ/ des premiers. Li Lincan, Zhang Kun *et al.* 1953:242 signalent une variante /ʂɣ́/ pour une syllabe prononcée /ʂó/ à l'intérieur d'une formule rituelle (formule qui représente, en un sens, une seule unité, dont chaque syllabe peut ne pas être perçue comme mot indépendant, et connaître des

changements phonétiques), confirmant l'affinité entre /ɣ/ et les voyelles d'arrière fermées et mi-fermées.

Dans les emprunts récents au chinois, les /u/ sont réinterprétés comme des /ɣ/ : ainsi, la négation bu^4 不 est prononcé /pɣ/ (s'agissant de l'initiale, rappelons que les b, d, g de la notation romanisée standard du chinois, le *pinyin*, notent des consonnes non voisées) ; « brouillard », wu^4 雾, est prononcé /ɣ́/.

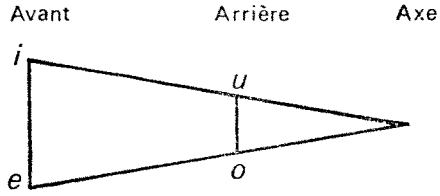
2.3. Notation des voyelles d'aperture moyenne

Le naxi ne compte que trois degrés d'aperture vocalique (comme il a été mentionné section 1.2, dans l'analyse des oppositions entre coronales rétroflexes et non rétroflexes devant /o/). Les voyelles cardinales étant arrangées en quatre degrés, un choix doit être fait pour noter les voyelles d'aperture médiane. Aucun argument phonologique décisif ne rapproche ces voyelles des voyelles fermées ni des voyelles ouvertes. Au plan phonétique, les réalisations fluctuent ; les notations de Mazaudon et Michailovsky 1979 varient entre une notation mi-fermée /e/ et une notation mi-ouverte /ɛ/ pour la voyelle d'avant ; pour la voyelle d'arrière non arrondie, c'est /ʌ/ (voyelle mi-ouverte) qui est choisi.

Notre choix repose en premier lieu sur un argument phonétique : la voyelle d'arrière arrondie se situe plutôt dans la région de [o] que de [ɔ], la voyelle d'avant non arrondie dans la région de [e] plutôt que de [ɛ], comme en témoignent les figures 3.46 à 3.55 (pp. 475-479). La voyelle /o/ se réalise [ø] en l'absence de consonne initiale, et après /h/ (et, dans le dialecte de DYZ, après toutes les initiales)[1], ce qui fournit un argument supplémentaire en faveur de la notation /o/ : si l'on adopte la notation comme voyelle mi-fermée, la variation allophonique consiste en une antériorisation, sans changement de degré d'ouverture. La notation des voyelles non arrondies d'arrière se calque sur celle des voyelles arrondies : /ɯ/ en regard de /u/, /ɤ/ en regard de /o/. Au plan phonétique, ce choix reflète de façon assez directe la réalisation du /ɤ/ : sur la figure 1.2, cette voyelle figure précisément aux côtés de /o/, fournissant argument pour sa notation comme /ɤ/. Le fait ne se vérifie pas systématiquement sur les triangles vocaliques (figures 3.46 à 3.55, pp. 475-479), du fait de la tendance à une réalisation antériorisée de /o/ (nette dans les données de F2). En revanche, /ɯ/ est souvent réalisée comme une voyelle centrale, se rapprochant de [ɨ] ; cette variation allophonique, de même que l'allophone [ø] de /o/, qui se généralise dans les dialectes de la plaine de Lijiang, tend à alléger l'espace des voyelles d'arrière, très fourni en naxi. André Martinet résume ainsi la tension entre symétrie du système et difficulté phonétique à maintenir distinctes les voyelles d'arrière : « ... même dans le champ vocalique, privilégié à cet égard, l'économie articulatoire évidente que représente le choix du même angle

[1] La réalisation [ø] n'est pas systématique : M4 prononce parfois [ho]. Le locuteur M2 (He Jiren), originaire de Yang^{4}Xi^1 漾西, prononce invariablement [o], quelle que soit l'initiale.

d'ouverture du maxillaire pour les voyelles d'avant et celles d'arrière doit entraîner un écartement plus considérable à l'avant et une « distance » acoustique plus grande entre [e] et [i] qu'entre [o] et [u].



Le choix du même angle d'ouverture n'est pas une vue de l'esprit, comme le montre l'étude des faits de diphtongaison. La tendance à l'équidistance se manifeste par la fréquence des systèmes où la série d'avant a plus d'unités que la série d'arrière, c'est-à-dire où un type d'économie (tendance à l'équidistance) l'emporte sur l'autre (identité des angles d'ouverture). Il y a donc là, pour tout système phonologique, une source possible de déséquilibre » (Martinet 1975:60). **Au plan articulatoire et acoustique, l'arrondissement est impliqué à la fois dans la réalisation des corrélations phonologiques d'arrondissement, et, de façon indirecte, dans l'opposition entre voyelles d'avant et d'arrière** : l'arrondissement des lèvres pour les voyelles postérieures a pour effet d'abaisser leur deuxième formant (F2), ce qui explique que dans de nombreuses langues (comme le français) dans lesquelles les voyelles d'arrière n'entrent pas dans une corrélation d'arrondissement (en français, /u/ ne s'oppose pas à /ɯ/, ni /o/ à /ɤ/), ces voyelles sont fortement arrondies. (Inversement, un arrondissement d'une telle ampleur pour /y/ nuirait à sa bonne reconnaissance : l'abaissement du troisième formant qui en résulterait serait tel que le troisième formant deviendrait plus bas que le second ; F2 et F3 ne se renforceraient alors plus mutuellement, or la proximité et le renforcement mutuel de F2 et F3 est la principale caractéristique acoustique du /y/ français.) **Le naxi, en utilisant le trait distinctif d'arrondissement pour les voyelles d'arrière comme pour les voyelles d'avant, privilégie l'économie phonologique, contre la simplicité phonétique.**

2.4. Réalisation phonétique des voyelles d'arrière non arrondies

Au plan phonétique, la réalisation de voyelles d'arrière non arrondies demande à ce qu'un second formant bas soit produit sans arrondissement. Cette exigence acoustique éclaire l'observation selon laquelle la voyelle /ɯ/ « donne un aspect carré à la bouche » (observation due à Yang Huandian 杨焕典 dans une communication à la 27e Conférence internationale sur les langues et la linguistique sino-tibétaine, Paris, 1994 : « Nouvelles réflexions sur les voyelles tendues et relâchées en naxi – en réponse au professeur Dai Qingxia » 再论关于纳西语中的紧松元音问题 – 兼答戴庆厦教授, et confirmées par nos propres données). **Ce geste peut s'interpréter comme un *dés-arrondissement***, dans la mesure où il se distingue de l'arrondissement aussi bien que de l'étirement des lèvres (tel qu'il s'observe par exemple pour /i/) : l'arrondissement ferait courir le risque d'une confusion

avec /u/ ; l'étirement aurait pour effet de raccourcir la cavité résonante en avant de la constriction (en termes de modélisation acoustique, le *tube* en avant de la constriction), ce qui élèverait les deux premiers formants, alors que /ɯ/, voyelle fermée, nécessite un F1 bas. (Les photos 1 à 13, pages 510-513, montrent le visage du locuteur M4 de face pendant des voyelles tenues ; par souci d'exhaustivité, les clichés correspondant aux voyelles autres que /u/, /i/ et /ɯ/ sont également présentés.)

2.5. Notation des voyelles ouvertes

L'opposition entre les deux unités vocaliques les plus ouvertes est notée par He Jiren et Jiang Zhuyi 1985 comme une opposition de /a/ à /ɑ/ ; d'autres (Li Lincan, Zhang Kun *et al.* 1953, et certains ouvrages non spécialisés, tels que Kou Bangping 1995) opposent /æ/ à /a/ ; Mazaudon et Michailovsky 1979 opposent /a/ à /ɒ/.

L'opposition de /æ/ à /a/ ne paraît pas adéquate au plan phonétique, car le second membre de l'opposition est une voyelle d'arrière. Au plan phonologique, en l'absence de voyelle mi-ouverte, les voyelles ouvertes sont libres d'occuper une partie de l'espace vocalique de [æ], [ɛ], [ʌ] et [ɔ] ; les observations phonétiques (représentées sur les figures 3.46 à 3.55, pp. 475-479) indiquent que le /a/ est en fait réalisé comme une voyelle très ouverte (plus ouverte que le /ɑ/), ce qui plaide pour une notation /a/ plutôt que /æ/.

Pour ce qui concerne le choix entre les notations /ɒ/ et /ɑ/ pour la voyelle d'arrière ouverte, l'une et l'autre notation sont défendables : cette voyelle est légèrement arrondie. **Le léger arrondissement qui s'observe sur cette voyelle ouverte d'arrière peut néanmoins s'expliquer comme un corrélat secondaire de son trait de postériorité** : les voyelles postérieures se distinguent par un second formant (F2) bas ; l'arrondissement a pour effet acoustique d'abaisser les formants, et notamment le second.

2.6. La question de la « voyelle neutre »

La voyelle qui est ici notée /ɤ/ est /ə/ dans la notation de plusieurs auteurs dont He Jiren et Jiang Zhuyi 1985, et Fang Guoyu et He Zhiwu 1995. Comme il a été brièvement exposé plus haut, la notation /ə/ ne paraît pas adéquate au plan phonétique ; de plus, cette voyelle se comporte comme une voyelle d'arrière fermée, se prononçant avec une initiale vélaire (« onset-filler ») lorsqu'elle forme syllabe à elle seule : [ɣɤ] (ainsi, « écoper » se dit /ɤ̌/, prononcé [ɣɤ̌]).

Une voyelle neutre distincte de /ɤ/ peut former syllabe en naxi : forme de *syllabe minimale*, /ə/, porteuse d'un ton, qui apparaît dans certains mots disyllabiques, par exemple /ə bē/ « l'an passé », /ə ŋī/ « hier », /ə ī/ « maintenant, de nos jours », et comme syllabe initiale de nombreux termes de parenté disyllabiques, par exemple /ə́ nā/ « grand'mère » (祖母), qui

calquent le a[1] 阿 du chinois ; la syllabe /ə́/ forme également à elle seule le morphème d'interrogation totale. Elle se réalise phonétiquement [ə], sans semi-voyelle initiale.

Le choix de noter cette voyelle comme une voyelle neutre rejoint l'intuition de He Xueguang (notre locuteur M3), qui a ajouté un symbole vocalique pour noter cette voyelle, refusant son identification avec /ɤ/, proposée par la romanisation officielle du naxi. Les faits qui paraissent décisifs sont les suivants :

1) **Dans certains cas, il est clair que ce /ə/ est le produit d'un phénomène de réduction syllabique**. Ainsi, le morphème qui signale un procès en cours, /t^{h}è/, est généralement simplifié en /ə/ ; la forme complète /t^{h}è/ demeure substituable (*modulo* les nuances stylistiques attendues).

2) **Au plan phonétique, la réalisation des /ə/ varie selon la voyelle de la syllabe principale dans les mots disyllabiques, et d'un locuteur à l'autre**. La variation est en partie dialectale : à FK, la réalisation des /ə/ est proche de [a][1]. La locutrice F1 emploie plutôt [ə] ; elle juge « lourde » (mais non fautive) la prononciation de son frère (M1), souvent proche de /ɑ/. Le /ə/ connaît divers degrés d'harmonie vocalique selon les locuteurs (par exemple : [è mē] pour /ə mē/ « mère », chez le locuteur M4). Mazaudon et Michailovsky 1979 notent généralement cette syllabe par leur *a* d'arrière, [ɒ], lorsque la syllabe principale a pour voyelle /ɯ/, /y/ ou /ɑ/ (noté /ɒ/), et [ʌ] lorsque la syllabe principale a pour voyelle /e/ ou /i/. Ce type de situation paraît typique d'une voyelle neutre, comme par exemple en maleng brô (langue viet-muong), qui possède entre autres des disyllabes à présyllabe simple vocalique ; la réalisation phonétique de l'élément présyllabique dépend de l'aperture de la voyelle principale : /əkaˀ/ se réalise [akaˀ], et /əsœk/ [usœk]. Cette tendance à l'harmonie vocalique n'engendre pas d'opposition pertinente (Ferlus 1997:58).

2.7. Attaque des voyelles syllabiques

Dans certaines langues, les syllabes sans consonne initiale prononcées à l'isolée commencent par une attaque *dure*, glottalisée : ainsi en allemand, et en vietnamien de Hanoi : « Syllables which are written with no initial consonant—that is, the first symbol is a vowel letter—actually begin with a glottal stop » (Thompson 1965:4). En vietnamien, le phénomène est systématique. En naxi, la situation est différente selon le degré d'aperture de

[1] Il va de soi que les phénomènes de réduction et la tendance à l'harmonie vocalique ne sont pas l'apanage de la voyelle syllabique /ə/. Tous les mots grammaticaux connaissent une tendance à l'affaiblissement phonétique : ainsi, la négation, /mɤ̃/, se rapproche phonétiquement d'une réalisation [mɑ̃] dans la séquence NEG. + verbe copule, /mɤ̃ wù/.

la voyelle : les voyelles fermées et mi-fermées sont précédées d'une semi-voyelle, comme le montre le tableau 16.

phonème-syllabe	prononciation	Lieu d'articulation de la semi-voyelle	notation de He Jiren et Jiang Zhuyi 1985
i	ji	Palatal	ʐi
y	jy	Labio-palatal	ʐu
ɯ	ɣɯ	Vélaire	ɣɯ
ɤ	ɣɤ	Vélaire	ɣə
u	wu	Labio-vélaire	u
o	ø	Labio-vélaire	o

Tableau 16. Réalisation phonétique des syllabes constituées d'une voyelle fermée ou mi-fermée.

Cette distribution correspond aux attentes : semi-voyelle labiale devant les voyelles arrondies /y/ et /u/, palatale devant les voyelles d'avant fermées /i/ et /y/, vélaire devant les voyelles d'arrière /u/, /ɯ/, /o/ et /ɤ/. (La fricative vélaire voisée [ɣ] se comporte donc, en naxi, comme une semi-voyelle.)

Nous n'avons pas trouvé d'élément d'explication de l'antériorisation de /o/ lorsque cette voyelle forme syllabe à elle seule (et lorsqu'elle suit un /h/, /ho/ se réalisant [hø]).

La rime /e/ n'apparaît pas seule ; une application mécanique du principe de minimisation du nombre d'unités postulées pourrait conduire à interpréter les syllabes prononcées [ə] comme la réalisation de /e/ en l'absence de consonne initiale (ce que fait Fu Maoji 1984:315) ; il nous semble néanmoins que [e] et [ə] ne sont pas des allophones plausibles[1]. (Pour plus de détails sur la syllabe /ə/, voir ci-dessus, section 2.6.)

Les voyelles /a/, /ɑ/, /ə˞/ sont précédées, parfois d'une attaque glottique (par exemple : [ˀa], [ˀə˞]), plus fréquemment d'une attaque douce[2]. La rime /ɣ̩/ se réalise simplement [ɣ].

He Jiren et Jiang Zhuyi 1985 notent correctement /u/ la syllabe prononcée [wu]. En revanche, les syllabes /i/, /y/, /ɯ/ et /ɤ/ sont notées avec deux symboles : /ʐi/, /ʐy/, /ɣɯ/, /ɣə/. La semi-voyelle phonétique qui les accompagne est interprétée comme une fricative alvéo-palatale pour i et y. Lors d'une discussion sur cette question, He Jiren nous a expliqué

[1] Au plan diachronique, la syllabe [ə] possède à l'évidence plusieurs origines en naxi. David Bradley, lors d'une communication au séminaire « Langues tibéto-birmanes », soulignait que dans les langues ni/yi, la syllabe /ə/ se trouve au croisement de plusieurs phénomènes, et que ses origines diachroniques comme ses fonctions synchroniques sont complexes.

[2] Dans ce dernier cas, He Jiren (communication personnelle) propose de noter avec une initiale glottale voisée : [ʱa], [ʱə˞]. D'après nos impressions auditives, il ne paraît pas indispensable d'utiliser un symbole consonantique ; la simple notation [a], [ə˞] nous paraît suffire.

que ce choix était principalement motivé par un argument phonologique : le choix de la notation /ʐ̩/ permet en effet de compléter la série des fricatives et affriquées alvéolo-palatales, qui comprenait déjà, dans la notation de ces auteurs, /tɕʰ/, /tɕ/, /dʑ/, /ndʑ/ et /ɕ/. En réalité, une fois cette série réanalysée comme allophones de consonnes vélaires (voir ci-dessus, section 1.3), cet argument de symétrie tombe. He Jiren s'est finalement rangé à ce point de vue, et à l'idée d'étendre aux autres voyelles fermées et mi-fermées la notation phonémique qu'il avait déjà adoptée pour /u/.

La syllabe /ɤ/ est relativement rare ; elle est illustrée par /ʂɤ/ « puiser, écoper de l'eau » (lao[1] 捞). He Jiren et Jiang Zhuyi 1985 et Fang Guoyu et He Zhiwu 1995:429 notent ce mot /ɣʂɤ/ ; en revanche, He Zhiwu 1987:5 avait déjà avancé l'idée selon laquelle la voyelle /ɤ/ (qu'il note comme un schwa, /ə/), lorsqu'elle forme une syllabe à elle seule, se réalise phonétiquement [ɣɤ]. Le comportement de /ɤ/, voyelle d'arrière mi-fermée, est donc parallèle à celui de /ɯ/, voyelle d'arrière fermée. Parmi les voyelles d'arrière fermées et mi-fermées, les non arrondies qui forment syllabe commencent par [ɣ], les arrondies qui forment syllabe commencent par [w].

Au plan phonétique, la semi-voyelle initiale est clairement audible. La partie semi-vocalique *(glide)* comporte un bruit de friction. Pour la segmentation, la forme du signal a été employée : le signal d'une syllabe telle que [ji] a typiquement une forme *en ampoule*, l'amplitude du signal étant plus faible pendant la partie semi-vocalique ; l'écoute du signal permet un découpage approximatif en fonction du degré de friction perçu.

En parole continue, une attaque glottique peut être utilisée pour détacher un mot ; l'attaque glottique est une variante intonative. De même qu'en français, les mots courts, qui risqueraient de poser des difficultés de compréhension, sont renforcés par une attaque glottique : « Je m'appelle ʔYves », « Je m'appelle ʔAnne », « c'est un ʔara [un perroquet ara] » (observation due à Michel Launey pour le français, qui rejoint les observations de Dilley et Shattuck-Hufnagel 1996 sur l'anglais).

2.8. Possibilité de fusion entre une syllabe sans consonne initiale et la syllabe qui précède

En parole continue, l'attaque douce des syllabes sans consonne initiale fait qu'elles s'enchaînent à la rime de la syllabe précédente, d'où la possibilité de fusions syllabiques. Le phénomène s'observe sous forme de réalisations phonétiques amalgamées dans les récits enregistrés : par exemple, /sɑ̄ ō/, « opium », qui en isolation se réalise [sɑ̄.ō], est réalisé [sɑ̃ø] dans le récit « Vampire » du locuteur M4 (position exacte sur l'enregistrement : à 159 secondes 60 centièmes), réalisation qui nous paraît donner l'impression d'un monosyllabe. Le même phénomène peut s'observer sous une forme lexicalisée, par exemple la particule

finale /sjɤ̄/, qui proviendrait de /sè/ (accompli) + /jɤ̄/ (particule finale affirmative)[1]. Ces processus de réduction sont fréquents dans les langues de la région : un exemple récent en mandarin est la création de beng[2] 甭 (« pas la peine de... ») par coalescence de bu[2] yong[4] 不用 (NEG. + « utiliser, utilité ») ; le caractère utilisé pour noter la syllabe résultante a été créé en superposant les caractères des deux syllabes télescopées[2].

Un exemple observé en parole spontané paraît particulièrement révélateur : le mot « jambe », /kʰɯ̄ ō/ dans le dialecte de AS (réalisé [kʰɯ̄ ø̄] ; rappelons que /o/ est réalisé [ø], et /ho/ réalisé [hø]), est prononcé [kʰø̄ː] par la locutrice F1. Ce mot est porteur de deux innovations phonologiques :

- il présente la voyelle [ø] après occlusive, ce qui ouvre la possibilité d'une opposition entre (par exemple) /kʰø̄/ et /kʰō/ (chez cette locutrice, le mot « son » (声音), /kʰō/, est réalisé [kʰō], de même que chez tous les autres locuteurs de AS)
- il introduit la quantité vocalique comme trait régulier, pouvant acquérir valeur distinctive.

Dans la suite de l'évolution, il est imaginable que la différence de quantité s'estompe, laissant uniquement une différence de qualité, [o] et [ø] cessant dès lors d'être deux allophones du phonème /o/, et acquérant statut de phonème, avec pour première paire minimale : /kʰō/ (« son », 声音 ; homophone : « chapeau de paille », dou[3]li[4] 斗笠) et /kʰø̄/ « jambe ». Une autre possibilité serait que seule la différence de longueur soit conservée : « jambe » deviendrait alors /kʰōː/, et formerait une paire minimale avec /kʰō/.

Il n'est guère probable que la prononciation de la locutrice F1 acquière une influence aussi large ; elle n'emploie plus guère le naxi que lors de visites à ses grand'parents, et a fait le choix de parler exclusivement le chinois mandarin avec son fils. Les phénomènes de cet ordre n'en sont pas moins porteurs de changements latents. Si d'aventure cette locutrice effectuait un séjour prolongé dans son village d'origine, peut-être rétablirait-elle la prononciation disyllabique du mot « jambe » ; il n'est pas exclu que ce soit au contraire sa prononciation qui soit imitée, cette enfant du pays étant forte d'un succès universitaire et

---

[1] Le nombre des exemples est plus restreint à AS, dialecte sur lequel se concentre l'analyse synchronique présentée ici, que dans le dialecte de DYZ : ainsi, /kjɤ́/ « argent » (forme du dialecte de DYZ) provient très certainement de /ki ´ jɤ̄/ (même sens ; forme en usage dans le dialecte de AS).

[2] La tendance à la réduction à un monosyllabe d'unités lexicales composées de deux syllabes paraît être une tendance très ancienne dans les langues de la région (voir en particulier l'hypothèse d'une monosyllabisation du chinois archaïque au chinois ancien, formulée par Ferlus 1998). Elle peut aboutir à la création de schémas syllabiques atypiques : ainsi, en khmu, les mots /klɲaʔ/ « résine » et /klmeʔ/ « canne à sucre » constituent des exceptions au patron syllabique observé partout ailleurs dans la langue, qui n'admet pas les présyllabes du type /kl/ ; ces mots proviennent en réalité de disyllabes, /kə̆l.ɲaʔ/ et /kə̆l.meʔ/, dans lesquels /kə̆l/ signifie « arbre » (vietnamien *cây*) (Ferlus 2001b:13).

professionnel qui en fait *(to coin a phrase)* « The glass of fashion and the mould of form, / The observed of all observers » (*Hamlet*, 3.1).

2.9. Semi-voyelles

L'intuition de certains auteurs chinois (He Jiren et Jiang Zhuyi 1985, Fu Maoji 1981) est que les semi-voyelles auraient été entièrement absentes dans des états plus anciens du naxi. De fait, **un certain nombre de rimes comprenant une semi-voyelle sont des emprunts ; mais cela n'est pas vrai de toutes**. Le tableau 17 présente les finales complexes relevées par He Jiren et Jiang Zhuyi 1985 et indique lesquelles sont présentes dans nos données.

rime	exemple d'emprunts chinois	prononciation actuelle en mandarin standard	présence dans des mots « natifs » (ou emprunts bien intégrés)	probabilité d'un emprunt depuis la 2e moitié du XXe s.
/je/	/tsjĕ˧/ 节 « fête », /tʰjè/ 铁 « fer »	jie^{2}, tie^{3}	non	quasi-certitude
/ja/	/mjá/ 面 « farine ; nouilles », /ljà/ 粮 « céréale »	mian^{4}, liang^{2}	non	quasi-certitude
/jɑ/	/pjā/ 表 « tableau, table... » ; /pʰjá/ 漂 « flotter »	biao^{3} ; piao^{4}	oui : /jā kò/ « maison »	présomption
/jo/	/kʰi⁻ ljŏ˧/ 侵略 « envahir »	qin^{1}lue^{4}	*marginal* ; /kʰjó lō/ « fourmi » (M3)	quasi-certitude
/jɤ/			oui : /mjɤ̀/ « œil »	faible
/jə/			oui : kʰūī ŋgjə̄ « boiteux »	faible
/ɥe/	/tɥé/ 对 « correct, juste », /sɥē/ 官 « mandarin »[1]	dui^{4} ; guan^{1}	variante expressive de /y/, dans : /zō kʰɥè˧/ « la gent masculine », /mi˧ kʰɥè˧/ « la gent féminine »	forte présomption
/ɥa/	/hɥà/ 危险 « dangereux », /kʰɥá/ 劝 « exhorter »	xian^{3}, quan^{4}	non	forte présomption
/we/	/hwé/ 会 « réunion », /kwé/ 贵 « précieux », /wè/ 围 « encercler »	hui^{4}, gui^{4}, wei^{2}	non	quasi-certitude
/wa/			oui : /wā/ « gauche » (左)	assez faible
/wɑ/			oui : /wà/ : verbe copule	faible
/wɤ/			oui : /ŋgwɤ̄/ « se promener »	faible

Tableau 17. Illustration des rimes à semi-voyelles du naxi de AS, emprunts chinois inclus.

Les rimes à semi-voyelle du chinois mandarin qui ne figurent pas ci-dessus sont simplifiées lors de l'emprunt en naxi. Ainsi, guo^{4} 过 (dans le sens « célébrer (une fête) : 过节, et dans :

[1] Guo Dalie, He Jiren *et al.* 1999:66 proposent d'analyser le mot /sɥē/ « mandarin » (guan^{1} 官) comme une contraction de /sūī/ + /pʰē/, mais sans fournir de glose de ces deux dernières syllabes.

guo[4] yu[2] 过于 « trop, outre mesure », surtout employé dans une tournure négative) est simplifié en /kó/.

Dans le dialecte de FK, les réalisations [we] et [wɤ] alternent librement. Le choix de l'un ou de l'autre est affaire de préférence personnelle : le principal consultant (M12) prononce [wɤ], et corrige en [wɤ] la prononciation [we] du consultant M15.

Le degré de probabilité d'un emprunt des sept premiers groupes à date récente (depuis la seconde moitié du XXe siècle) est très élevé. Dans l'écriture syllabique geba (哥巴文), seules les quatre dernières rimes, /wɑ/, /wa/, /jɤ/ et /wɤ/, possèdent une notation (Fang Guoyu et He Zhiwu 1995:479-481), ce qui fournit un argument de leur présence dans la langue depuis une date relativement ancienne. La plus marginale est /wɤ/.

Il est imaginable, quoique non certain, que la rime /wɑ/ provienne de la diphtongaison d'un ancien /o/. Fu Maoji 1981 reconnaît les deux rimes /o/ et /wɑ/, mais les deux catégories ne se recoupent pas exactement avec celles du dialecte de AS : « haut, élevé » est /ʂò/ chez Fu Maoji 1981, /ʂwà/ dans nos données ; « extrêmement » est /zo̱/ chez Fu Maoji, /ʐwɑ̄/ dans nos données. L'avancement du travail de reconstruction n'est pas suffisant pour se prononcer sur les évolutions qui ont conduit à cette situation. De même, il est imaginable que la rime /wɤ/ provienne d'anciens /u/. Les seules rimes /wɤ/ attestées apparaissent après des vélaires ; l'association privilégiée des articulations vélaires et labiales, bien établie à travers les langues, pourrait expliquer qu'un passage de /u/ à /wɤ/ ait eu lieu après les vélaires à l'exclusion des autres ordres.

Signalons que le locuteur M4 oppose entre elles des syllabes /ku/ et /kwɤ/ que d'autres locuteurs (dont un frère de M4) jugent être homophones, et en variation libre, [kwɤ] ou [ku]. Ainsi, d'après M4, « gingembre » se dit /kwɤ̄/, et ne saurait être prononcé /kū/ ; le classificateur des cigarettes est /kwɤ́/ et non /kú/ ; l'interjection pour faire rentrer les vaches dans l'enclos est /kwɤ̄/ ; « rouler, enrouler » est /kwɤ̀/ ; tous mots qui ont une voyelle /u/ dans le dialecte de DYZ, et chez d'autres locuteurs de AS. Le /ku/ de ce même locuteur M4 est illustré par /kú bē/ « bouche » (zui[3] ba[1] 嘴巴). Il pourrait s'agir simplement d'une modification idiosyncratique, propre à ce locuteur. Elle suggère néanmoins la plausibilité d'un passage de /u/ à /wɤ/ sur certains mots à date relativement récente[1].

Comme il a été mentionné en section 2.1, **la solution provisoirement retenue consiste à reconnaître l'existence de /w/ et /j/ comme semi-voyelles médiale**, jugeant que le naxi possède une structure syllabique (C)(G)V, où C représente une consonne, G (pour *Glide*)

[1] Le locuteur M4 a également tendance à diphtonguer le /y/ en /yᵊ/ lorsqu'il se trouve en position finale de groupe intonatif ou lorsque la syllabe reçoit une insistance particulière. Ce phénomène est particulièrement saillant dans la liste de nombres suivis du mot « montagne », /ŋgỳ/, élicitée pour étudier la modification des tons des chiffres et nombres en fonction du ton de la syllabe suivante : il prononce [duū.ŋgỳᵊ, ni ˉŋgỳᵊ...].

une semi-voyelle /w/ ou /j/, et V une voyelle, seule composant indispensable de la syllabe. La distribution des semi-voyelles médiales est très limitée : l'une et l'autre n'apparaissent que devant /ɑ/, /a/, /ɤ/ et /ə/, et après certaines consonnes seulement. **Cette analyse revient à diminuer de huit unités** (wɑ, wa, wɤ, wə, jɑ, ja, jɤ, jə) **l'inventaire des unités vocaliques**, au prix de la reconnaissance d'une structure syllabique plus complexe que la simple association d'une initiale et d'une unité vocalique. Ce choix permet également de réanalyser comme /wə/ la combinaison [və] (marginalement présente dans le lexique, voir section 1.4.3), réanalyse qui ne présente toutefois pas un degré complet de certitude.

2.10. La voyelle rhotique : confirmation de l'absence d'oppositions de qualité de voix en naxi

Les diverses descriptions publiées du naxi ne mentionnent pas d'usage distinctif de la qualité de voix (Fu Maoji 1940 ; Li Lincan, Zhang Kun *et al.* 1953 ; Rock 1963-1972 ; Mazaudon et Michailovsky 1979 ; Fu Maoji 1984 ; Fang Guoyu et He Zhiwu 1995 ; Pinson 1996, 1998 ; et tout récemment la description du naxi oriental, ou mosuo, par Liberty Lidz, en préparation). L'ouvrage de He Jiren et Jiang Zhuyi 1985 indique explicitement l'absence de ce trait distinctif[1], par ailleurs présent dans toutes les langues du groupe yi, relativement proches parentes du naxi. **Il ne serait pas utile de s'étendre sur le fait s'il n'avait été remis en question par Yang Huandian** (杨焕典), dans un article paru en mars 1984 (« Discussions sur le système phonémique du naxi/ 论纳西语音位系统 »). Celui-ci se fonde sur l'idée selon laquelle le naxi est une langue du sous-groupe ni/yi, et souligne l'invraisemblance de l'absence en naxi d'un trait présent dans toutes les autres langues yi. Il défend l'idée selon laquelle le naxi aurait lui aussi une opposition entre voyelles tendues et relâchées.

L'idée selon laquelle le naxi serait une langue du sous-groupe yi est en réalité dépassée depuis Bradley 1975 : les ressemblances de surface entre le naxi et les langues yi ne manquent pas, mais les correspondances phonétiques sont peu régulières ; le naxi serait à placer dans une branche voisine des langues yi, plutôt qu'à l'intérieur du groupe. A la lumière de cette réanalyse, l'absence en naxi d'une caractéristique commune aux langues yi paraît moins surprenante. Au plan des observations synchroniques, Dai Qingxia (戴庆厦) signale son désaccord avec le point de vue de Yang Huandian dans un article intitulé « A propos de la question des voyelles tendues et relâchées du naxi » (paru dans la revue chinoise *Minzu Yuwen* : "关于纳西语的松紧元音问题" , 民族语文1993年1期, pp. 27-31 et 36). En réponse, Yang Huandian attaque violemment Dai Qingxia dans une

[1] Citation originale : "元音松紧不分音位/ Le caractère tendu ou relâché des voyelles n'est pas employé pour distinguer les phonèmes" (He Jiren et Jiang Zhuyi 1985:15).

communication à la 27e Conférence internationale sur les langues et la linguistique sino-tibétaine (Paris, 1994) : « Nouvelles réflexions sur les voyelles tendues et relâchées en naxi – en réponse au professeur Dai Qingxia 再论关于纳西语中的紧松元音问题 – 兼答戴庆厦教授 ») : fort de l'autorité que lui confère son statut de locuteur natif du naxi, il soutient que la voyelle /ə̰/ serait en fait un /z̩/ (noté / ɿ /, suivant en cela la proposition de Chao Yuen-ren pour le chinois) auquel s'ajouterait un trait laryngal de *tension*. La position de la langue serait la même pour le [ɿ] allophone de /ɯ/ (par exemple dans /zɯ̄/ « herbe ») et pour le /ə̰/ (par exemple dans /bə̰̄/ « hôte, invité »).

Les mesures de formants réalisées ici confirment l'idée traditionnelle selon laquelle la voyelle /ə̰/ est nettement distinguée des autres voyelles par son articulation orale, reflétée par la position des formants. Sa caractéristique principale est un troisième formant extrêmement bas. Au plan glottique, les mesures de quotient ouvert montrent un recouvrement entre la plage de variation de la rime /ə̰/ et celle des autres rimes, absence de différence qui indique qu'au plan laryngal, /ə̰/ ne se singularise pas. Ces observations confirment que, dans les dialectes de NL, DYZ, AS et FK, les voyelles ne comportent pas de spécification de qualité de voix.

3. Prononciation des phonèmes de la variété locale de chinois

Ce volet de l'enquête, d'orientation synchronique, présente un intérêt indirect pour l'étude diachronique. L'étude de la prononciation des phonèmes de la variété locale de chinois par les locuteurs naxi permet en effet d'aborder le domaine du contact entre langues sans se limiter aux seuls emprunts lexicaux. Les contacts entre langues ont joué un rôle considérable dans l'évolution des langues tibéto-birmanes : l'influence constante du chinois (directe et indirecte) et celle du tibétain, qui lui-même contient de nombreuses couches d'emprunts au chinois, créent des situations de convergence aréale. L'étude des changements dus à l'influence du chinois peut contribuer à une meilleure compréhension des processus d'échange entre langues et, partant, de l'histoire des langues de la région.

3.1. Remarques générales

3.1.1. Situation actuelle de la langue par rapport au chinois

La génération des plus de 40 ans connaît principalement la variété locale de chinois : un dialecte du Sichuan, parlé par des colons venus du Sichuan à une époque relativement récente (quelques indications historiques sont fournies par Guo Dalie et He Zhiwu 1999). La diffusion rapide de la télévision met actuellement les Naxi de toutes générations en contact avec le mandarin standard.

Dès l'école primaire (obligatoire à partir de sept ans), les jeunes Naxi sont scolarisés en chinois mandarin standard ; les jeunes locuteurs paraissent avoir tendance à utiliser les mêmes sons lorsqu'ils parlent naxi, ce qui rejoint les observations d'André Martinet citées plus haut : le système phonologique de la langue maternelle et celui de la langue nationale restent distincts ; une modification phonétique peut avoir lieu dans les sons qui se ressemblent dans les deux langues en contact (Martinet 1956:53). Au vu de l'influence actuelle du chinois sur le naxi (qui s'oriente dans le sens d'un remplacement progressif du naxi par le chinois), il paraissait utile d'étendre l'enquête linguistique à la deuxième langue des consultants naxi.

Dans les termes du référentiel proposé par Brenzinger, Dwyer *et al.* 2003 (*Language Vitality and Endangerment*), la situation des dialectes naxi étudiés peut se décrire comme suit :

	critère	notation du naxi, et commentaire
1	transmission	4 (*unsafe*). Tous les enfants comprennent la langue, mais certains (à la ville en particulier) ne la pratiquent pas, et n'acquièrent pas la compétence à s'exprimer dans la langue.
2	nombre de membres de l'ethnie	élevé : près de trois cent mille personnes sont officiellement Naxi.
3	proportion de locuteurs parmi les gens appartenant au groupe ethnique	4 (*unsafe*) : une grande partie des locuteurs ; une faible proportion des enfants issus de mariages mixtes
4	sphères d'emploi	3 *(dwindling domains)* : la langue est employée dans la sphère domestique, mais la langue dominante commence à l'y concurrencer.
5	place dans les nouveaux domaines technologiques et les média	0-1 : les mots référant à des réalités technologiques nouvelles sont empruntés, sans création de néologisme ; la langue des média est exclusivement le chinois.
6	matériaux disponibles pour l'enseignement de la langue	0 : quelques manuels ont été créés dans les années 1980, mais aucun n'a été distribué et n'est employé en pratique
7	attitude du gouvernement	2 : assimilation active. Une éducation en chinois est offerte ; la pratique du naxi n'est pas encouragée.
8	attitude des locuteurs	1 : seuls quelques individus sont favorables à la préservation de la langue et y contribuent, d'autres sont indifférents ou favorables à sa disparition.

*Tableau 18. Situation du naxi dans les termes du référentiel proposé par Brenzinger, Dwyer et al. 2003 (*Language Vitality and Endangerment*)*

Le sentiment de beaucoup de consultants naxi (notamment les plus âgés) est que le naxi d'aujourd'hui est déjà une langue mélangée, emprunts chinois et mots plus anciens étant inextricablement mêlés.

Avant la présentation d'observations sur les emprunts chinois en naxi, le rappel de quelques observations générales, puis un survol de deux travaux publiés au sujet de l'influence du chinois mandarin sur d'autres parlers, permettront de cadrer le volet phonétique de cette question.

Boyd Michailovsky (séminaire LACITO) observe que les vecteurs du changement sont les *bons bilingues*, ceux qui prononcent avec précision la deuxième langue (langue d'influence),

tandis que les locuteurs peu à l'aise dans leur deuxième langue calquent dans celle-ci les structures de leur langue maternelle, qui pour leur part demeurent bien en place. Le même constat figure déjà chez Martinet 1975:45-46 :

> ...la contagion ne joue guère là où le bilinguisme a permis d'assimiler un nombre considérable d'emprunts. Dans ces emprunts, la prononciation, défectueuse à l'origine, finit par s'améliorer. Mais la mutation que représente cette amélioration ne reste pas limitée aux emprunts ; elle s'étend à des mots du vocabulaire indigène. Si, dans un mot d'emprunt, /y/ est tout d'abord réalisé comme [u] et si, par la suite, les sujets parlants parviennent à corriger leur [u] en [y] dans ce mot, certains /u/ de mots indigènes pourront suivre.

Ce constat est vérifié expérimentalement par Peng Shuhui 1993 dans une étude de la prononciation des fricatives du mandarin prononcées par des apprenants dont la langue maternelle est l'amoy. Le comportement des apprenants permet de distinguer trois groupes :

1) Les apprenants peu avancés prononcent les /f/ et /χ/ du mandarin comme le /hw/ (h labialisé) et le /h/ de l'amoy. Lorsqu'ils parlent amoy, leur prononciation reste inchangée par leur apprentissage du mandarin.

2) Les apprenants relativement avancés apprennent le /f/, et prononcent de façon correcte ce son absent de leur langue maternelle (leur prononciation est proche de celle de locuteurs natifs du mandarin) ; leur /χ/ demeure mal prononcé : il est proche du /h/ amoy, dont il n'est pas clairement distingué par les locuteurs. Lorsqu'ils parlent amoy, leur prononciation reste inchangée par leur apprentissage du mandarin.

3) Les apprenants très avancés prononcent /f/ et /χ/ de façon proche de la réalisation par les locuteurs natifs du mandarin. Lorsqu'ils parlent amoy, leur prononciation n'est plus canonique : elle se rapproche du [χ] mandarin, ou s'y identifie tout-à-fait.

Par analogie avec cette étude, il paraît possible de prédire que le /h/ naxi, dans les positions où il se réalise comme [h] (c'est-à-dire devant /e/, /a/, /o/, /ɚ/) a toutes chances de se renforcer en [x] ou [χ] à mesure que le bilinguisme avec le mandarin progressera ; et que les consonnes vélaires palatalisées du naxi, qui à l'heure actuelle sont moins fortement palatalisées que leurs homologues en mandarin, s'identifieront tout-à-fait à celles-ci.

Au-delà de l'adoption de nouvelles articulations, et du remplacement d'articulations plus anciennes par celles de la langue influençante, les emprunts sont vecteurs de changements structurels. Parmi les travaux réalisés par le CRLAO (Centre de Recherches Linguistiques sur l'Asie Orientale), en partenariat avec l'Académie des sciences humaines de Chine, sur les emprunts chinois dans les autres langues de Chine, Sagart et Xu Shixuan 2001 proposent une étude des emprunts chinois en hani (langue du groupe ni/yi, branche ni/yi-birmane du tibéto-birman). Le hani possède trois tons, définis par leur registre : Haut, Moyen, Bas. Les auteurs distinguent deux principales périodes dans les emprunts. Au cours de la période M1 (jusque vers la fin de la dynastie Qing, en 1911), les syllabes empruntées se plient à la

phonotactique du hani. La seconde période, M2, commence à peu près avec la période de la République de Chine ; elle correspond à des emprunts plus directs, au mandarin de Lüchun, localité qui, jugent les auteurs, a beaucoup grandi en importance au cours de la période. Des syllabes au ton montant, calqué du mandarin, commencent à apparaître. Après 1949 (couche d'emprunts baptisée « M2b »), le bilinguisme se développe, et l'intégration des emprunts devient de plus en plus faible. « Later M2b borrowings (...) basically do not mind contravening Hani phonotactics » (p. 37).

La situation est globalement la même en naxi : les couches anciennes d'emprunt au chinois sont intégrées au système phonémique, et ne sont pas immédiatement identifiables ; les emprunts récents sont en fait des mots chinois faiblement intégrés qui remplacent peu à peu les mots plus anciens.

3.2. La langue chinoise parlée par l'informateur M4

Une enquête détaillée sur la prononciation de la variété locale de chinois parlée par le locuteur M4 a été réalisée. Elle permet de connaître le système phonologique qui concurrence celui du naxi dans l'expérience linguistique de la génération née autour de 1950.

Le locuteur M4 a appris le chinois à l'école primaire ; enseignement assuré par un enseignant naxi pratiquant la variété locale de chinois. L'écriture phonétique du mandarin standard n'était enseignée (par un enseignant ne sachant pas la prononcer avec précision) qu'à partir de la quatrième année d'école primaire, et il n'a pu suivre cet enseignement.

Ce locuteur connaît également un peu de mandarin standard (il écoute la télévision le soir après dîner, pendant une heure environ), et est conscient de la variété des prononciations du chinois : différence entre le mandarin standard et le chinois des colons venus du Sichuan, installés dans le voisinage, auprès de qui il a appris la langue ; différence entre les parlers du Sichuan et celui du Hunan, constatée lorsqu'il a été témoin de difficultés de compréhension entre locuteurs de différents dialectes chinois ; et différence entre le chinois des locuteurs natifs et sa propre prononciation.

3.2.1. Maintien intermittent d'une distance phonétique entre naxi et chinois

Dans l'ensemble, la structure syllabique et l'inventaire phonémique du chinois parlé par le consultant M4 sont très proches de celles du naxi. La rime –uo est simplifiée en –o, rime qui se prononce [ø] après /h/ et lorsqu'elle forme syllabe à elle seule, suivant en cela la règle qui s'applique à l'ensemble des combinaisons /o/ et /ho/ du naxi de AS. L'attaque des voyelles syllabiques est douce en « chinois langue seconde » comme en naxi. Les coronales sont réalisées rétroflexes devant les voyelles d'arrière mi-fermées et fermées : ainsi, /lú/

« chemin » (路 lu[4]) est réalisé [l̥ú]. Les tons sont les mêmes qu'en naxi : un ton haut ; un ton moyen ; un ton bas ; un ton montant. Il paraît donc clair que le système tonal de la langue seconde est le même que celui de la langue maternelle ; le consultant peut indiquer les tons des mots chinois en utilisant la méthode dont nous sommes convenus pour les mots naxi, ce qui confirme que le système est le même.

Néanmoins, **la prononciation du « chinois langue seconde » se signale par un effort conscient du locuteur** : celui-ci cherche à rejoindre un modèle de prononciation imparfaitement maîtrisé mais qu'il sait différent de celui de sa langue maternelle. Cet effort se traduit par les particularités phonétiques suivantes :

- la nasalisation sporadique des voyelles et l'ajout sporadique d'une consonne nasale finale, d'une façon qui n'est pas corrélée avec les finales nasales que demanderait une prononciation canonique des mots chinois concernés : ainsi, 盖子 gai[4] zi « couvercle » est prononcé [kán.tsɯ̄] (alors que le mot ne comporte de finale nasale ni dans les dialectes mandarins du sud-ouest, ni en mandarin standard)
- la diphtongaison de /ɤ/ en [ɤ[w]], par exemple dans gou[1] 沟 (水沟) « rigole, canal », prononcé [kɤ̄] ou [kɤ̄[w]] d'une fois l'autre (prononciation de cette rime en mandarin du sud-ouest : /əu/ (Yang Shih-Feng 1984, p. ex. page 1224)
- la réalisation de la fricative vélaire [χ] devant /e/ (ainsi, guo[3] he[2] 果核 « noyau de fruit », est prononcé [kō χè] plutôt que [kō hè]), alors que dans les mots naxi, la combinaison [χè] n'existe pas : l'allophone de /h/ devant /e/ est simplement [h] : par exemple, /hē mē/, « lune », se réalise [hē.mē]
- la réalisation des rimes à semi-voyelles comme des séquences de voyelles, /ja/ se rapprochant de [i.a] (par exemple dans la prononciation de yan[1] 烟 « cigarette »).

Ces particularités n'ont pas valeur distinctive : en « chinois langue seconde » pas plus que dans les dialectes mandarins du sud-ouest il n'existe de rime [ɤ] distincte de [ɤ[w]], ni de syllabe [he] distincte de [χe] ; les caractéristiques phonétiques en question ont pour effet de signaler qu'il s'agit de mots chinois. Ces observations rejoignent un constat déjà effectué par Trubetzkoy :

> **Les éléments distinctifs anormaux :** En dehors du système phonologique normal, beaucoup de langues présentent en outre des éléments phonologiques particuliers, qui apparaissent avec des fonctions tout à fait spéciales. A cette catégorie appartiennent avant tout les « sons étrangers », c'est-à-dire les phonèmes qui sont empruntés au système phonologique d'une langue étrangère (...). ... de tels « sons étrangers », une fois introduits dans la langue, ne sont pas toujours employés à leur « juste place » : ils sont une caractéristique d'origine étrangère et peuvent par conséquent apparaître dans un mot senti comme étranger, qu'ils soient à leur place dans ce mot ou non. Ainsi le mot étranger Telephon est à Vienne très souvent prononcé avec une voyelle

> nasale (telefõ), et les Tchèques remplacent k par g dans des mots étrangers comme plakat, balkon, etc. (Trubetzkoy 1939 [1976]:245)

3.2.2. Suppression des nasales finales, et phénomènes d'hyper-correction

Les nasales finales sont fréquentes et assurent un grand nombre de distinctions en chinois mandarin ; les locuteurs naxi ressentent le besoin de les réaliser lorsqu'ils s'expriment en chinois ; comme par ailleurs ils éprouvent la plus grande difficulté à se souvenir de ces consonnes finales (absentes de leur langue maternelle) même lorsqu'ils sont parvenus à en maîtriser la prononciation, leurs efforts aboutissent à des hyper-corrections. Ainsi, le locuteur M5 ajoute une nasale vélaire au nom de l'ethnie Yi (en chinois : yi^{2} zu^{2} 彝族) : il le prononce ying^{2} zu^{2} (prononciation que l'on pourrait rendre, en chinois, par les caractères 莹族). Cette erreur rappelle (entre cent autres situations semblables) l'hyper-correction des locuteurs du dialecte chinois de Qingdao (qing^{1}dao^{3} 青岛) : leur dialecte ne connaît pas la distinction entre les rimes –ai et –ei du mandarin standard, confondues en –ei ; entendant des finales –ai en mandarin standard, parler de prestige, eux aussi se piquent de prononcer des finales –ai ; mais n'ayant pas appris la distribution des –ai et des –ei lors de l'acquisition du lexique de leur langue maternelle, ils remplacent leurs –ei par des –ai de façon systématique ou aléatoire, ce qui conduit à des erreurs : par exemple, Pékin (bei^{3} jing^{1} 北京) sera prononcé bai^{3} jing^{1}, ce qui vaut à ces infortunés le sobriquet de bai^{3} jing^{1} ren^{2} 百京人 (qui pourrait se traduire par : « paikinois »).

A mesure qu'un locuteur parle mieux chinois, il en vient à réaliser les nasales finales ; ainsi, M3 parle un chinois mandarin très courant, qui comprend une réalisation correcte des nasales finales ; les mots chinois à nasale finale qui lui viennent lorsqu'il parle naxi ne perdent pas leur nasale finale lorsqu'ils se glissent dans son discours.

3.2.3. Conservation des oppositions portées par les nasales lorsqu'elles sont en position initiale

Certaines syllabes à nasale finale sans consonne initiale du mandarin standard sont prononcées avec une nasale initiale dans les dialectes chinois locaux, et empruntés tels quels en « mandarin langue seconde » par les locuteurs naxi : ainsi, « tranquille » 安静 an^{1} jing^{4} est prononcé /ŋā ki ´/. Cette prononciation reflète celle du dialecte chinois avec laquelle les locuteurs naxi sont en relation de voisinage : dans le Sichuan, 安 se dit /ŋanˀ/ ou /ŋãˀ/ selon les dialectes (Yang Shih-Feng 1983:1235, Li Rong 1998:274). De même, 爱 ai^{4} est prononcé /ŋai/ en mandarin du Sichuan(avec un ton descendant-montant ; Li Rong 1998:175) ; cette syllabe est prononcée par notre locuteur naxi (M4) comme /ŋá/ (par exemple dans /ljà´.ŋá/ « se lier d'amour » : 恋爱 lian^{4} ai^{4}). De façon similaire, 棉袄 mian^{2} ao^{3} (« « veste ouatée, vêtement rembourré ») est prononcé /mjà.ŋū/ (prononciation de la deuxième syllabe dans le

dialecte mandarin du Sichuan : /ŋau/ [Li Rong 1998:220]), 便宜 pian^{2} yi^{4} « bon marché » devient /p^{h}já.ni˴/, non par l'effet d'une resyllabation par les locuteurs naxi, mais du fait que la syllabe 宜 se prononce /ɲi/ dans la variété locale de chinois (Li Rong 1998:7).

3.2.4. Changements dont le « chinois deuxième langue » est potentiellement porteur

3.2.4.1. Changements dans les initiales : dans les mots de « chinois deuxième langue », /v/ a statut d'initiale, la combinaison /ni/ est très fréquente, /ŋ/ peut former syllabe

Les syllabes à initiale /w/ du mandarin sont prononcées avec un /v/ initial en « chinois deuxième langue », ce qui fait entrer ce phonème (qui jusque-là apparaissait exclusivement comme rime, /ụ/) dans la catégorie des initiales.

L'introduction d'un nombre croissant d'emprunts a pour effet de modifier la fréquence respective des unités concernées dans le lexique, effet qui est particulièrement sensible dans le cas de la combinaison /ni/. La variété locale de chinois possède pourtant un /ɲi/ aussi bien qu'un /ni/ ; mais les nasales coronales devant /i/ sont toutes réalisées non palatalisées en « chinois langue seconde » : [ni], par exemple dans [p^{h}já.ni˴] « bon marché » (便宜 pian^{2} yi^{4}), mot dont la deuxième syllabe a pourtant une initiale palatalisée en chinois du Sichuan. Nous n'avons pas à l'heure actuelle d'hypothèse au sujet de ces emprunts qui ne suivent pas le chemin qui paraissait le plus plausible au plan phonétique. (Rappelons que la syllabe /ni/ est également attestée en naxi, mais est beaucoup moins fréquente que la syllabe /ŋi/, réalisée [ɲi] : voir section 1.3.)

La première personne du singulier (我, prononcée wo^{3} en mandarin standard) est prononcée [ŋ̄] en « chinois langue seconde » de notre informateur[1] ; cette syllabe n'est pas attestée dans les mots naxi, où le plus proche équivalent est / ŋ̣ /, prononcé [ŋɱ] et non [ŋ]. (Cette forme est nettement distincte du pronom naxi de 1^{e} personne, /ŋɤ̈/.)

3.2.4.2. Changements dans les rimes : développement de rimes à semi-voyelle

Les rimes /ɥe/ et /ɥa/ sont fréquentes en « chinois langue seconde », ainsi que /jo/, dans / ljō ʂɣ̣ / « saule » (柳树 liu^{3} shu^{4}), et /we/, dans /t^{h}wē/ « jambe » (腿 tui^{3}).

Les semi-voyelles ont, pour l'oreille naxi, un *parfum de chinois*, du fait de leur nombre et de leur fréquence élevés en chinois par comparaison avec le naxi. Le locuteur M4 paraît avoir une prédilection pour la rime /ja/ (qui apparaît par exemple dans l'emprunt /mjá/ « nouilles », chinois mian^{4} tiao^{2} 面条), prédilection qui se traduit par son extension dans le lexique, mot par mot, d'une façon idiosyncratique : l'expression très fréquente /dūˍ.mə̀/,

[1] La prononciation du pronom de 1^{e} personne 我 dans la majorité des dialectes chinois du Sichuan est /ŋo/ ; dans certains dialectes, /o/, ou encore /ŋəu/ (Yang Shih-Feng 1984:1718).

« un peu », devient dans son parler /dɯ̄.mjà/ ; la particule finale /sjɤ̄/, très probablement issue de la fusion des particules /sè/ (accompli) et /jɤ̄/ (particule finale affirmative), devient fréquemment /sjā/. Le locuteur est conscient de ces spécificités, et connaît les prononciations /dɯ̄.mə̀ɤ/ et /sjɤ̄/, qui sont en usage à AS comme à DYZ.

Le statut particulier de cette rime /ja/ tient sans doute à sa très grande fréquence en « chinois langue seconde », environ quatre fois plus élevée qu'en « chinois parlé par des locuteurs natifs », puisque pour une oreille naxi, les rimes –ia, –ian, –iaŋ et –iao du chinois sont uniformément entendues –ia (le naxi ne connaissant pas de consonnes finales) ; un locuteur naxi qui écoute du chinois entend donc quantité de /ja/ qui se succèdent. (Dans le dialecte du Sichuan, ces rimes sont notées ia, iɛn, iaŋ, iau par Li Rong 1998:8).

Ces observations qui peuvent paraître anecdotiques rejoignent en fait des observations sur des langues variées, qui suggèrent que **des *phonèmes de prestige* peuvent s'étendre mot par mot dans le lexique, brouillant les correspondances attendues entre langues et dialectes** (voir notamment Martinet 1975:45-46). L'exemple classique, en français, est le mot *haut*, issu du latin *altus*, dont le /h/ initial paraît être dû à l'imitation (probablement phonétique, en tout cas orthographique) du mot correspondant dans des parlers germaniques. Michel Ferlus propose d'interpréter certains mots tibétains, dont la prononciation est aberrante au vu des évolutions phonétiques de cette langue, comme le fruit d'une modification d'un des phonèmes du mot, empruntant au chinois, non pas le mot correspondant, mais un de ses phonèmes (« ...the Tibetan word is the result of a compromise between an inherited form and an acquired segment »: Ferlus 2000, « On a mode of borrowing from Middle Chinese into Proto Tibetan: A new look at the problem of the relatinoship between Chinese and Tibetan »). Michel Ferlus explique par le même processus le passage du mot « riz, nourriture », /kɤm/ (*cơm* en orthographe vietnamienne) à /ɣɯɤm/ (*ɣươm* en orthographe vietnamienne) dans certains dialectes dits « hétérodoxes » : ces dialectes ne connaissaient pas les spirantes qu'apportait le parler vietnamien du nord lors de son extension vers le sud ; les locuteurs, par imitation de la variété prestigieuse de vietnamien, ont introduit ces articulations dans certains mots, d'une façon qui ne correspond pas à leur distribution dans la langue influençante, brouillant les correspondances attendues (Ferlus 2001a, « Les hypercorrections dans le thổ de Làng Lỡ, ou les pièges du comparatisme » ; document provisoire)[1].

---

[1] En synchronie, un phénomène du même ordre se manifeste en français dans des créations *plus anglaises que l'anglais* : expressions qui ne sont pas réellement anglaises, mais correspondent à la représentation que les Français se font de la langue anglaise. Le « pin's », un temps à la mode, se dit en fait *lapel badge* en anglais ; mais le monosyllabe « pin » est un mot bien plus commode pour un locuteur français que *lapel badge ;* la terminaison *'s* qui lui est ajoutée est perçue comme typiquement anglaise ; elle apparaît également dans le titre

3.2.4.3. Changements dans les tons

Il est bien établi que l'accentuation d'une langue étrangère est perçue d'après les cadres fournis par la langue maternelle ; une étude typologique a récemment été menée à ce sujet par Peperkamp et Dupoux 2002, « A typological study of stress 'deafness' ». Les remarques ci-dessous visent à éclairer indirectement l'espace tonal naxi, en mettant en lumière, par le traitement des emprunts chinois, les régions de *surdité tonale acquise*, qui correspondent aux régions de variation allophonique des tons naxi.

Fang Guoyu et He Zhiwu 1995:86 décrivent ainsi les équivalences tonales : « Les tons du naxi n'ont pas la même forme que les tons chinois, de sorte que les Naxi qui lisent des caractères chinois commettent régulièrement des erreurs. Par exemple, le ton *yang*^{2} *ping*^{2} 阳平 [ton montant] du pékinois devient ton bas descendant [ton 21 dans la notation de Chao Yuen-ren, dont la notation équivalente en « lettres-tons » est ˨˩, et qui est noté ainsi par de nombreux linguistes chinois], le ton *yin*^{1} *ping*^{2} 阴平 [ton décrit comme « haut », et noté 55, ou ˥ ; il n'est en fait pas au sommet de la plage du locuteur] devient ton moyen (33, ˧), le ton *qu*^{4} 去 [descendant] devient ton haut [55, ˥] ; l'ancien *ru*^{4} *sheng*^{1} 入声 [qui provient de syllabes à occlusives finales] est lu comme ton bas montant ˩ ; enfin, le ton *shang*^{4} 上 [qui est actuellement un ton bas à contour complexe en mandarin] est lu comme un ton moyen ˧. On ne peut comparer directement les tons des deux langues ; la façon dont les Naxi lisent les caractères chinois permet de dégager des règles de correspondance, qui restent néanmoins assez grossières. » (Ces mêmes indications figurent également dans la courte grammaire de He Zhiwu 1987:8.) Le tableau 19 résume cette correspondance.

numéro du ton correspondant en mandarin actuel ; terme traditionnel ; et description de la réalisation standard actuelle	ton naxi utilisé
1 : yin^{1} *ping*^{2} 阴平 [ton décrit comme « haut », et noté 55, ou ˥ ; il n'est en fait pas au sommet de la plage du locuteur]	M / ˉ/
2 : yang^{2} *ping*^{2} 阳平 [ton montant]	L / ˋ/
3 : *shang*^{4} *sheng*^{1} 上声 [ton bas à contour complexe : long, descendant, avec remontée en forme isolée]	M / ˉ/
4 : *qu*^{4} *sheng*^{1} 去声 [descendant]	H / ˊ/
ancien *ru*^{4} *sheng*^{1} 入声 [qui provient de syllabes à occlusives finales]	LH /ˋˊ/

Tableau 19. Traitement des tons des mots chinois lors de l'emprunt en naxi, d'après Fang Guoyu et He Zhiwu 1995:86.

Les équivalences les plus claires sont celles entre le ton 4 du mandarin et le ton H du naxi, entre le ton 1 du mandarin et le ton M du naxi. L'identification est totale chez les locuteurs naxi. **Le ton descendant du chinois est entendu comme haut car le ton haut naxi possède un**

---

de la revue française *Top's Cars*, autre faux anglicisme qui n'est pas correct en anglais. Une différence avec les faits naxi est que dans ces exemples français, l'emploi du *'s* n'est pas étendu à des mots indigènes.

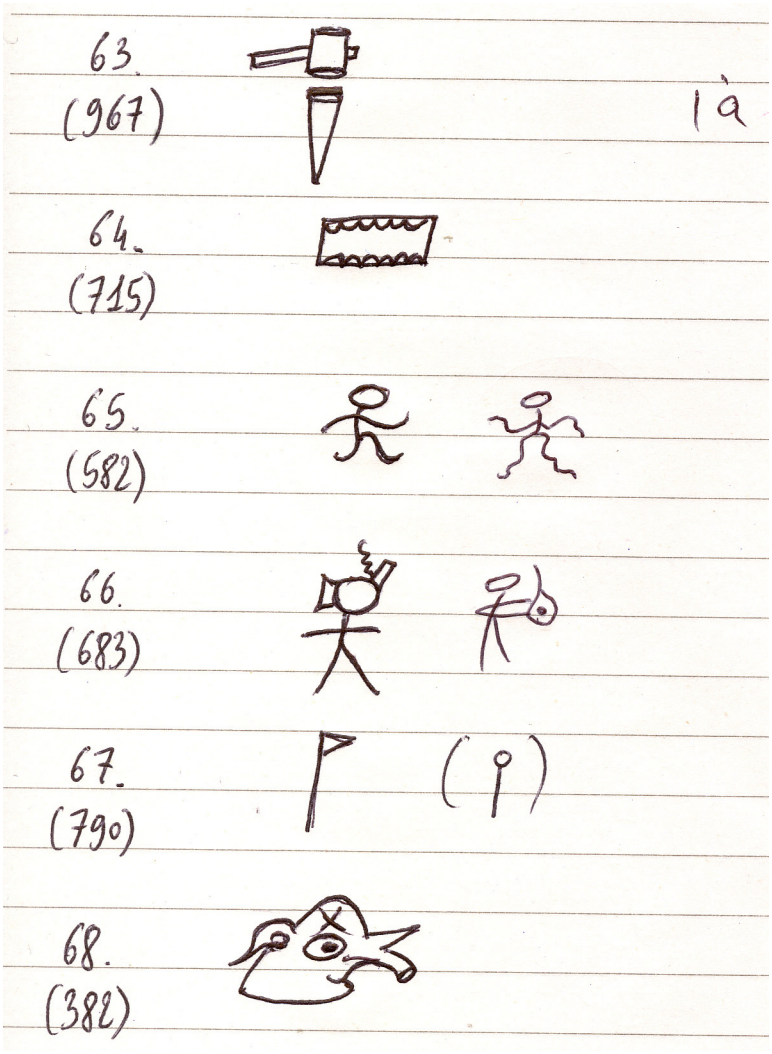
allophone descendant (l'expression *allotone* est parfois employée dans les études tonales) : cet allophone apparaît lors de la mise en valeur locale d'une syllabe ; la descente est particulièrement marquée lorsque cette syllabe se trouve en outre en position finale de groupe. **Les locuteurs du naxi ne retiennent donc du ton descendant chinois que son point de départ très élevé** ; ce point de départ est plus élevé que celui du ton 1 du mandarin, ton communément décrit comme haut, mais qui l'est moins que le ton haut du naxi.

Ce fait met en lumière une caractéristique phonétique des tons du chinois mandarin standard que leur notation dans les manuels (d'orientation phonologique) pourrait avoir tendance à faire oublier. Le ton 1 du mandarin est noté ^{55} (reprenant les conventions de notation proposées par Chao Yuen-ren)[1], donc comme un ton haut. Le point de départ du ton 4 du mandarin (ton descendant) est également décrit comme niveau maximum (le ton 4 est décrit comme ^{51}). Cette notation est reprise d'ouvrage en ouvrage (par exemple, tout récemment, dans la description du chinois dans la revue de l'Association Phonétique Internationale de juin 2003 : Lee Wai-Sum et Eric Zee 2003). L'interprétation des tons 1 et 4 du mandarin par les locuteurs naxi rappelle qu'au plan phonétique, le ton 1 serait plutôt ^{44} ou ^{33} que ^{55} : ton plat de niveau moyen ou moyen/haut, mais non au point le plus élevé du registre du locuteur. Il s'agit certes d'une remarque phonétique et non phonologique : un modèle simplifié à l'extrême ne nécessite que deux registres, haut et bas, le ton 1 étant haut, le ton 2 bas + haut, le ton 3 bas, le ton 4 haut + bas. Pour autant, ces faits phonétiques méritent d'être rappelés : dans la perspective de l'enseignement, l'apprenant n'est pas nécessairement familier de la distinction entre niveau tonémique et niveau phonétique, et serait peut-être plus efficacement guidé par des manuels qui décriraient le ton 1 comme « moyen/haut » et le ton 4 comme « descendant rapidement de très haut » que par une notation par ^{55} et ^{51}, respectivement. Dans une perspective de recherche, la modélisation du ton 1 comme un ton /Haut/ paraît certes très défendable ; pour autant, il ne s'agit pas nécessairement du fin mot de la question.

Un exemple d'usage des diacritiques tonals du chinois mandarin par un locuteur naxi illustre l'équivalence entre ton 4 mandarin et ton H naxi. Le locuteur M8 a accepté de participer à

[1] Dans cette notation, le premier chiffre indique le point de départ de la courbe mélodique et le dernier son point d'arrivée, sur une échelle de 1 (niveau le plus bas) à 5 (niveau le plus haut). Une notation équivalente est ˥ : la barre verticale offre un point de repère ; le tracé horizontal permet de visualiser le contour du ton ; le contour ainsi dessiné est strictement équivalent à celui que décrivent les chiffres ^{55}. Ce système de notation, qui remonte à Chao Yuen-ren 1930 (« A system of tone letters »), est dit « tonémique » par son auteur (le terme anglais employé est « toneme », p. 24). Quant à la réalisation (« tone-value », *loc. cit.*), elle peut être notée à droite de la barre verticale : par exemple, une réalisation quelque peu descendante du ton haut pourra être notée ˥˦. Dans l'opposition qui a actuellement cours entre « phonétique » et « phonologie », la notation à gauche de la barre verticale sera dite *phonologique*, celle à droite de la barre étant d'orientation *phonétique*.

l'expérience de lecture de mots naxi dans deux phrases-cadre différentes (expérience rapportée au chapitre III) ; il connaît l'écriture pictographique traditionnelle, non la romanisation du naxi ; la liste des mots à lire lui a donc été présentée sous forme de pictogrammes. Lors de la séance préparatoire préalable à l'enregistrement, nous passons les mots en revue, pour vérifier que sa lecture coïncide bien avec le mot désiré ; dans les cas où le caractère possède plusieurs lectures, le consultant ajoute spontanément une annotation qui lui permette de se souvenir de la syllabe qu'il lui est demandé de prononcer. Dans ce travail de *bricolage*, le consultant fait flèche de tout bois : parfois il substitue un caractère homophone qui ne possède qu'une lecture ; parfois il ajoute une notation en écriture geba (écriture phonétique traditionnelle) ; enfin, dans un cas, il utilise une notation de la notation romanisée *(pinyin)* du chinois, qu'il était en train d'apprendre à la date de l'enquête : il s'agit du caractère « frapper », le premier sur le document scanné ci-dessous (n°63 de la liste de mots ; le numéro 967 renvoie au numéro assigné à ce caractère dans le dictionnaire de Fang Guoyu et He Zhiwu 1995). (L'explication de cet exemple est singulièrement compliquée par le fait que les mêmes diacritiques ont un sens tout différent en *pinyin* et en alphabet phonétique international : **en pinyin, *là* transcrit la syllabe /la/ porteuse d'un ton *descendant*, le ton 4**.)



Le caractère en question peut se lire /lá/ (« frapper un grand coup ») aussi bien que /tȳ/ (« heurter à une porte, donner un petit coup sur quelque chose ») ; pour refléter la prononciation /lá/, le consultant choisit une notation en pinyin qui présente le ton 4 : dans cette équivalence, il choisit le ton descendant du chinois, non son ton haut. Cet exemple offre une confirmation en synchronie de l'équivalence entre ton 4 du mandarin (haut-descendant) et ton H du naxi, tandis que le ton 1 du mandarin (ton plat, souvent décrit comme « ton haut », mais qui en fait n'est pas réalisé au sommet de la plage de fréquence fondamentale) correspond à un ton M en naxi.

Certains flottements constatés dans les tons des mots empruntés au chinois tiennent au fait que les locuteurs naxi sont familiers à divers degrés avec plusieurs dialectes chinois ; ceux qui connaissent la réalisation canonique du ton s'appliquent plus ou moins à le réaliser, selon la situation d'énonciation. Ainsi, sur le disque compact d'un concert du musicien naxi Xuan Ke (宣科), celui-ci présente le morceau de musique 浪淘沙 en le prononçant, à plusieurs reprises, comme /lǎŋ.tʰàw.ʃā/, réalisant les tons selon l'usage de la variété locale de chinois mandarin[1] ; en revanche, au terme de son introduction, pour annoncer solennellement au public qu'il va maintenant entendre ce morceau de musique, il redit son titre en détachant les syllabes, prononçant avec soin ses tons selon la forme canonique du mandarin standard : lang^{4} tao^{2} sha^{1}. (Xuan Ke 1999 ; 浪淘沙 *Lang Tao Sha* est le deuxième morceau ; le passage cité se trouve au début de la plage 2 du disque.)

La réinterprétation des tons chinois selon les catégories déjà présentes en naxi ne signifie pas que les emprunts soient sans effet sur l'évolution du système tonal naxi. Le principal changement dont les emprunts chinois sont vecteurs est le développement du ton montant lexical ; il paraît probable qu'avant l'emprunt de mots chinois à ton montant, les contours montants n'existaient en naxi que comme produits de processus morpho-phonologiques, lexicalisés sur de rares mots lexicaux[2] ; les emprunts chinois accroissent nettement la présence du ton lexical montant dans le lexique : dans la liste d'environ 1000 caractères élicitée auprès du consultant M4, environ 15% sont prononcés avec un ton montant /˨˦/[3].

---

[1] Notre notation des tons de cet exemple est approximative. Le ton de la première syllabe est montant, celui de la 2^{e} syllabe bas, le troisième moyen/haut.

[2] Ces phénomènes sont décrits dans une communication de colloque (Michaud 2003) et dans un article soumis pour publication.

[3] Dans sept cas, deux tons différents ont été prononcés à des moments différents (l'enquête a eu lieu en trois temps : élicitation sans enregistrement, premier enregistrement, deuxième enregistrement), et le consultant a indiqué que les deux formes étaient également possibles.

Ces observations n'épuisent pas le domaine complexe de l'étude des emprunts chinois en naxi et de leur influence sur le système phonologique de cette langue. Les influences sont anciennes, nombreuses et variées, comme le notent Fang Guoyu et He Zhiwu 1995, He Jiren et Jiang Zhuyi 1985, Guo Dalie et He Zhiwu 1999 : dans la région de Lijiang même, le naxi est de longue date en contact avec les dialectes des colons (venus principalement du Sichuan) et des soldats cantonnés sur place (Lijiang a été une importante ville de garnison au début de la dynastie Ming [明朝, 1368-1644], le zèle des seigneurs féodaux de Lijiang ayant par la suite dispensé le suzerain de poster ses propres troupes dans la ville), ainsi qu'avec la langue officielle employée par les administrateurs.

Brève note sur les changements syntaxiques

Les changements syntaxiques induits par le contact avec le chinois ne sont pas étudiés en détail ici. Signalons simplement certains télescopages d'expressions chinoises et naxi : la formule chinoise hao^{3} le^{0} « 好了 ! » « bon ! / c'est bon ! / ça y est ! » est calquée en /hò˝ lɤ̄/, ajoutant la particule chinoise le^{0} 了 au mot naxi /hò˝/, qui signifie « correct, bien, convenable »[1]. (La tournure en *bon naxi* serait : /t^{h}ɑ́ sè/.) De même, des télescopages ont lieu à la faveur de l'homophonie entre la particule possessive /gɤ̄/ du naxi et le classificateur générique ge$^{4/0}$ 个 du chinois : « celui-ci », zhe^{4} ge^{0} 这个, est calqué en : /ʈʂ^{h}ū kɤ́/, formule composée du déictique naxi /ʈʂ^{h}ū/ et du classificateur calqué du chinois ge$^{4/0}$ 个.

4. Phénomènes à l'échelle du mot : fusion syllabique et intégration nominale

4.1. Fusion syllabique

Parmi les créations de monosyllabes par fusion de disyllabes, certaines semblent le fruit de processus productifs : ainsi, le morphème d'accompli /sè/ entre en coalescence avec le verbe dans les formes suivantes :

/bū sè/ > /bɤ̀/ (/bū/ : « aller », temps présent ou futur)

/hū sè/ > /hɤ̀/ (/hū/ : « aller », temps passé)

/lū sè/ > lɤ̀/ (/lū/ : « venir »)

/ts^{h}ù sè/ > /ts^{h}ɤ̀/ (/ts^{h}ù/ : « arriver »)

La forme contractée a le même sens que la forme complète, *modulo* les différences stylistiques attendues (He Jiren et Jiang Zhuyi 1985:52).

---

[1] Le mot /hò˝/ n'est pas perçu par les locuteurs comme un emprunt ; s'il s'agissait d'un emprunt, ce serait très certainement he^{2} 合 (comme le suggère He Zhiwu 1987:8), non hao^{3} 好.

La fusion de syllabes crée de nouvelles combinaisons ; la recherche de l'économie phonologique tend à ramener ces combinaisons nouvelles aux combinaisons existantes. Ainsi, la particule finale /sjɤ̄/, née de la fusion de l'accompli /sè/ et de la particule affirmative /jɤ̄/ (/sjɤ̄/ et /sè jɤ̄/ demeurant interchangeables en synchronie), est parfois prononcée [çjɤ̄] en parole continue (ce qui correspond, au plan phonémique, à /hjɤ̄/ : voir ci-dessus, section 1.4.3), notamment par les locuteurs M2 et M3, bien que chez ces mêmes locuteurs la forme canonique (indiquée lorsque le mot est demandé isolément) demeure /sjɤ̄/. Cette variabilité paraît être en partie due à l'effet palatalisant de la rime /jɤ/, qui partage cette propriété avec les rimes /y/ et /i/ : ainsi, /dɯ̄ sỳ wà/, « c'est pareil », est parfois prononcé [dɯ̄ çỳ wà] par M3 (ce qui correspond, au plan phonémique, à /dɯ̄ çỳ wà/, et est noté ainsi par Mazaudon et Michailovsky 1979:1) ; /si¯/, « pauvre », est prononcé /çi¯/ par leur consultant L. Outre cet effet phonétique qui tend à palataliser l'initiale, la tension vers un passage de /sjɤ̄/ à /hjɤ̄/ tient sans doute également à la meilleure intégration de cette dernière syllabe au système : /sjɤ/ est un *hapax*, la rime /jɤ/ n'apparaissant après aucune fricative ni affriquée autre que /h/ (combinaison qui, elle, est bien attestée, par exemple dans /hjɤ̀/, « se reposer », réalisé comme [çjɤ̀], et qui ne peut se réaliser [sjɤ̀]).

4.2. Phénomènes d'intégration nominale

En naxi comme en anglais, la composition nominale se fait de façon très simple : par juxtaposition, sans marque morphologique qui signale la nominalisation d'un verbe ou inversement l'emploi d'un nom comme prédicat. Dans d'autres langues monosyllabiques, la composition nominale s'accompagne de changements réguliers (voir par exemple la description par Hartmann-So 1985 des changements de voyelles et de tons lors de la composition nominale en daai chin, langue tibéto-birmane parlée en Malaisie : « Morphophonemic changes in Daai Chin »). **En naxi, les phénomènes d'*intégration nominale* (qui correspond à ce que Saussure 1916 et Frei 1929 nomment « agglutination ») ne présentent pas la même régularité : ils sont seulement sporadiques**, et concernent, selon les cas, le ton ou la voyelle.

4.2.1. Modifications tonales

4.2.1.1. Exposition des faits

Dans un certains nombre de mots disyllabiques, la première syllabe porte un ton lexical moyen, alors que le même lexème porte un ton bas lorsqu'il apparaît comme monosyllabe (formant mot à lui seul).

A l'échelle du mot disyllabique naxi, toutes les combinaisons tonales sont pourtant attestées : HH, HM, HL, MH, MM, ML, etc. En synchronie, le naxi est donc clairement une langue du type que Matisoff nomme « omnisyllabique » (possédant un ton par syllabe). Certains mots (mots lexicaux, ou *mots phonologiques* composés de deux morphèmes) portent néanmoins des tons qui ne correspondent pas à ceux que possèdent, à l'isolée, les monosyllabes dont ils sont composés, comme l'illustre le tableau 20.

mot de base, au TON L	sens du mot de base	deuxième syllabe	sens de la deuxième syllabe	mot composé (1e syll. au TON M)	sens du composé
nv̩̀	pois	hə̀	vert	**nv̩̄ hə̀**	pois vert, petit pois
gi`	eau	ʂù	(non établi)	**gi¯ ʂù**	loutre
gi`	eau	tsè	utiliser	**gi¯tsè pá**	bassine pour se laver le visage
gi`	eau	ŋgò	(non établi)	**gi¯ ŋgò**	inondation
gi`	eau	hò	(non établi)	**gi¯ hò**	rivière, canal
ʂù	fer	k^hū	tige *(identité à confirmer)*	**ʂū k^hū**	trépied sur le foyer (pour y poser casserole ou bouilloire), 三角
ʂù	fer	ndzò	pont	**ʂū ndzò**	pont de fer
mjɤ̀	œil	nù	noir	**mjɤ̄ nù**	égoïste
gi`	maison	t^hò	être adossé à	**gi¯ t^hò**	l'arrière de la maison, le terrain derrière la maison
gi`	maison	kā	devant	**gi¯ kā**	le devant de la maison, le terrain en face de la maison

Tableau 20. Exemples de disyllabes dont la première syllabe ne porte pas le ton qui est le sien lorsqu'elle forme mot à elle seule (c'est-à-dire sous forme monosyllabique).

Dans ces exemples, la séquence qui serait attendue, par application du processus de composition par simple juxtaposition qui est très courant en naxi, est LL ou LM, selon les cas ; la séquence observée est ML ou MM, selon les cas.

« Loutre d'eau » (shui^3 ta^3 水獭), /gi¯ ʂù/, comporte manifestement la syllabe /gi`/ « eau » et une syllabe /ʂù/ qui se réfère plus particulièrement à l'animal ; certains locuteurs (dont M3) désignent d'ailleurs la loutre par la seule syllabe /ʂù/. De même, « tique » (水蛭 shui^3 zhi^1), /pý/, est /gi¯ pý/ chez le locuteur M2. « Empereur » se dit /k^hà/ (emprunt probable au parler des envahisseurs mongols de la dynastie des Yuan 元, 1271-1368 : « Khan ») ; « la capitale » (lieu de résidence de l'empereur) est /k^hā dỳ/ (Fang Guoyu et He Zhiwu 1995:121), où les locuteurs reconnaissent, de façon transparente, le mot « empereur » et le mot /dỳ/ « lieu, terre ». (Le locuteur M1, jeune homme qui à la date de l'enquête poursuivait des études universitaires à Pékin, prononce /k^hà dỳ/, sans le changement de ton.) La modification tonale est particulièrement fréquente devant les adjectifs de couleur : « dent »

/dzà/, « dent blanche » /dzā pʰə̀/ (/pʰə̀/ : « blanc ») ; /ki ̀/ « nuage », /ki ̄ pʰə̀/ « nuage blanc » ; /à/ « roc », /ā pʰə̀/ « roc blanc », /ā nà/ « roc noir », /ā hỳ/ « roc rouge », /ā hə̀/ « roc vert ». « Œil » se dit /mjɤ̀/, « le blanc de l'œil » est /mjɤ̄ pʰə̀/, « la prunelle » /mjɤ̄ nà/ (Fang Guoyu et He Zhiwu 1995:387). « Soja » (huang^{2} dou^{4} 黄豆) se dit /nỳ/, « petit pois, pois vert » (lü^{4} dou^{4} 绿豆) se dit /nȳ hə̀/ (/hə̀/ : « vert »).

Le processus est sporadique, non systématique : ainsi, /ŋgỳ/ « montagne » donne /ŋgỳ nà/ « montagne noire », non /ŋgȳ nà/. Dans quelques cas, un ton moyen sur la syllabe initiale de mot devient un ton haut : ainsi, « terre » se dit /ʈʂɯ̄/, et « terre blanche » /ʈʂɯ́ pʰə̀/, « terre noire » /ʈʂɯ́ nà/, « terre rouge » /ʈʂɯ́ hỳ/, « terre jaune » /ʈʂɯ́ ʂù/. Le mot ainsi formé ne constitue pas un élément lexical autonome, au sens spécialisé (comme « peau-rouge » ou « pied-noir » en français), mais véhicule simplement la notation de couleur que l'adjectif a à l'origine. Par exemple, « terre blanche » veut dire : de la terre blanche, et pas (par exemple) une qualité spécifique d'argile ou de calcaire.

Pour que le changement tonal ait lieu, il n'est pas indispensable que le sens de chaque élément se soit estompé : le changement a eu lieu dans des composés comme /ʂū ndzò/, « pont de fer », qui pourtant demeurent transparents aux locuteurs (« fer » + « pont ») ; de même, /mjɤ̄ nà/ « égoïste » est analysé sans difficulté par les locuteurs en /mjɤ̀/ « œil » et /nà/ « noir ».

Le changement n'a jamais été observé avant un ton H : ainsi, « clou » se dit /ʂù pú/, mot dans lequel apparaît le mot /ʂù/ « fer », sans changement tonal, à la différence des deux exemples du tableau 20.

D'une façon qui paraît similaire, un changement catégoriel systématique (obligatoire) s'observe dans le cas des pronoms de première et deuxième personne suivis d'une particule au ton M **/nɯ̄/ ou /gɤ̄/**[1] (/nɯ̄/ est une particule marquant l'*origine*, l'une de ses valeurs étant

[1] Si marginal que paraisse le phénomène, il pourrait avoir une importance dans l'évolution de la langue. Le passage de /ŋɤ̀/ et /nù/ à /ŋɤ̄/ et /nɯ̄/ dans les constructions possessives pourrait être le prélude à la création de formes possessives à part entière, par ellipse de la particule /gɤ̄/ rendue inutile au plan fonctionnel par la différence entre /ŋɤ̀/ et /ŋɤ̄/. La marge de variation que conserve (au plan phonétique) le pronom personnel /ŋɤ̀/ tend à prévenir cette émancipation de la forme modifiée /ŋɤ̄/ qui en est issue ; plusieurs exemples rencontrés en parole continue (dans des récits) suggèrent pourtant que cette évolution est imaginable. Ainsi, dans le récit « Vampire » raconté par le locuteur M4, une veuve qui s'en revenait du marché est dévorée par une truie-vampire ; le monstre enfile ses vêtements et se fait passer pour elle à sa maison afin de dévorer ses deux filles ; à plusieurs reprises au cours du récit, le vampire s'adresse aux deux enfants en les appelant « mes filles », /ŋɤ̄ mi ́/. Cette forme synthétique, qui véhicule une proximité intime, n'est pas remplaçable, dans le contexte du récit en question, par /ŋɤ̄ gɤ̄ mi ́/, expression bien formée mais analytique plutôt que synthétique. Ce phénomène rappelle, en anglais, l'opposition entre la construction possessive en *'s* et celle en *of* (que Henry Adamczewski et Claude Delmas décrivent comme une opposition entre *phase 2* et *phase 1*, P. Cotte en

d'expliciter le statut d'*agent* que possède un actant ; **/gɤ̄/** sert à former la construction possessive), comme le montre le tableau 21.

	Pronom	**Pronom suivi de la particule /nɯ̄/**	**Pronom suivi du possessif /gɤ̄/**
1e personne	**ŋɤ̀**	**ŋɤ̄** nɯ̄	**ŋɤ̄** gɤ̄
2e personne	**nɯ̀**	**nɯ̄** nɯ̄	**nɯ̄** gɤ̄
3e personne	$t^hɯ̄$	$t^hɯ̄$ nɯ̄	$t^hɯ̄$ gɤ̄

Tableau 21. Modification du ton des pronoms de première et deuxième personne lorsqu'ils sont suivis de la particule marquant l'origine ou de la particule de possessif.

Le changement de ton qui a lieu dans ces constructions ne concerne que les pronoms personnels ; ce n'est donc pas un effet des particules /nɯ̄/ et /gɤ̄/, qui en elles-mêmes sont sans influence sur le ton de la syllabe qui précède : ainsi, /$ts^hɯ̀$/ « fantôme » donne, après ajout de la particule marquant l'agent : /$ts^hɯ̀$ nɯ̄/, non /$ts^hɯ̄$ nɯ̄/. De même, /dɯ̀/ « grand » ne devient pas /dɯ̄/ dans la construction suivante :

(1) ŋɤ̀ dɯ̀ gɤ̄ mɤ̄ ŋi⁻.
1e sg. grand POSS. NEG. vouloir
« Je n'en veux pas de grand. » (contexte : lors du choix entre plusieurs vanneries)

Enfin, dans le cas des mêmes pronoms lorsqu'ils sont suivis du déictique-topicalisateur /$t^hɯ̄$/, deux formes existent de façon concurrente : un mot phonologique dont chacun des tons est identique à celui des morphèmes dont il est composé, et un mot phonologique dont l'un des tons n'est pas identique à celui d'un des morphèmes dont il est composé. Les combinaisons /ŋɤ̀/ + /$t^hɯ̄$/ (1e pers. sg. + DÉICT./TOP.) et /nɯ̀/ + /$t^hɯ̄$/ (2e pers. sg. + DEICT./TOP.) peuvent se réaliser /ŋɤ̄ $t^hɯ̄$/ et /nɯ̄ thɯ̄/ aussi bien que /ŋɤ̀ $t^hɯ̄$/ et /nɯ̀ $t^hɯ̄$/ (la manipulation a été tentée à partir de la forme /ŋɤ̄ $t^hɯ̄$/ présente dans un récit du locuteur M4).

4.2.1.2. Hypothèses sur l'origine diachronique du changement tonal

Les changements décrits ci-dessus ont pour effet l'établissement d'un *schéma descendant* à l'échelle du mot (par exemple : MM plutôt que LM, ML plutôt que LL…), mais dans certains cas, la différence de ton entre éléments et composé est inverse : /$ʈʂ^hù$/, « perle », donne /$ʈʂ^hù$ ná/ « perle noire » (Fang Guoyu et He Zhiwu 1995:137), alors que « noir » se

---

recourant à la notion de *nominalisation ;* voir Adamczewski et Delmas 1982, Cotte 1998). Dans le même récit apparaît également la tournure /ŋɤ̄ là ʈʂɤ́/, « mon poignet » (/là ʈʂɤ́/ : « poignet »), dans laquelle /ŋɤ̄/ joue le rôle de possessif ; dans ce contexte, il n'est pas possible de le remplacer par /ŋɤ̀/.

dit /nù/, avec un ton L. Certaines de ces modifications tonales pourraient remonter à des oppositions précédant l'apparition des tons dans leur forme actuelle, qui auraient conditionné le passage de certaines syllabes de mots disyllabiques (par exemple la deuxième syllabe de l'actuel /ʈʂʰù ná/) dans une catégorie différente du monosyllabe correspondant. D'autres langues posent des problèmes similaires, qui sont encore non résolus (en lahu : Matisoff 1973a ; en lalo : Björverud 1998 ; en tamang, « il existe des glissements de ton en composition, mais les règles restent obscures » [Mazaudon 1973:34]).

Martine Mazaudon (communication personnelle) suggère que **le naxi a pu connaître, au cours de son histoire, une phase pendant laquelle il aurait entamé une évolution vers un système à *ton de mot*** (système dont le tamang fournit une illustration exemplaire : voir Mazaudon 1973, et la section 3.1.5.4 de la Discussion [= chapitre IV]). Au cours de cette hypothétique période, **un certain nombre de mots phonologiques en seraient venus à porter un unique ton**, au lieu d'un ton par syllabe ; neutralisation partielle des contrastes tonals, dont le « ton léger » du mandarin de Pékin (évoqué en section 3.2.2.1 de la Discussion [= chapitre IV]) fournit un exemple observable en synchronie. Il est phonétiquement plausible, s'agissant de tons ponctuels, que leur réalisation à l'échelle d'un disyllabe plutôt que d'un monosyllabe soit quelque peu *descendante* au plan phonétique : telle est effectivement la tendance qui s'observe en synchronie en naxi, la courbe de fréquence fondamentale des tons H, M et L (réalisés sur des monosyllabes en phrase-cadre) étant légèrement descendante (voir par exemple les figures 3.31 à 3.45 ; les tracés d'un autre locuteurs, publiés dans Michaud 2003a, montrent la même tendance). Le début de transition vers un système prosodique *à ton de mot* aurait d'abord affecté certains mots lexicaux courants, et aurait affecté plus systématiquement certains composés dont les éléments avaient un faible degré d'autonomie : tel est le cas des composés réduplqués disyllabiques, si l'on se fie à l'hypothèse (qui nous paraît probable) selon laquelle ils proviendraient à l'origine d'un simple redoublement à l'identique ; tel est également le cas des séquences *pronom + particule grammaticale* du tableau 21. Un tel composé peut, plus facilement qu'un autre mot prosodique disyllabique, voir le ton d'une de ses syllabes s'effacer au profit d'un *ton de mot*. Pour poursuivre cette hypothèse diachronique, il faut imaginer que **l'évolution ainsi entamée aurait ensuite été contrecarrée**, par des facteurs non encore déterminés (peut-être la pression de langues « omnisyllabiques » voisines). Chaque syllabe des disyllabes qui étaient passés au *ton de mot* se serait alors vu réassigner un ton propre ; ce processus se serait fait par une réinterprétation dans le cadre proposé par le paradigme des trois principaux tons lexicaux H, M et L ; c'est alors que **la différence phonétique de hauteur entre les deux syllabes des mots phonologiques en question aurait donné lieu à une réinterprétation** : par exemple, /gî ʂù/ serait un moment passé par un stade où le mot ne portait qu'un unique ton L, dont la cible était atteinte tardivement ; sa réalisation phonétique,

proche de [gɪ̄ ʂù], aurait ensuite été réinterprétée comme telle au plan phonologique (d'où l'actuel /gɪ̄ ʂù/) lors du « retour à l'omnisyllabisme ».

Le tableau 22 présente, d'une façon qui s'inspire des recherches de Jacqueline Vaissière (Vaissière 1983, « Language-independent prosodic features »), une courbe stylisée de la tendance phonétique au passage d'une valeur de F_0 élevée à une valeur basse, tendance constatée à l'échelle du mot (et d'unités prosodiques plus larges). La pente de la courbe est plus forte en début et en fin. Ce tableau vise à aider à se représenter le passage diachronique d'un disyllabe dont chaque syllabe possède un ton, puis à un stade où le mot porte un seul ton (réalisé, au plan phonétique, par une courbe de fréquence fondamentale légèrement descendante), et enfin un stade où la partie de courbe portée par chacune des syllabes est réinterprétée indépendamment, *réinterprétation monosyllabique* aboutissant aux schémas actuels.

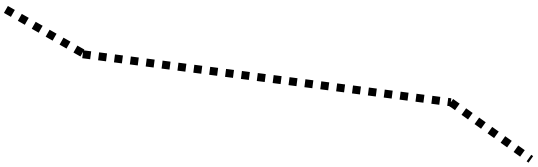
	disyllabe dont chaque syllabe a un ton		stade (hypothétique) du ton de mot	« réinterprétation monosyllabique »	
Niveau tonal	1e syll.	2e syll.	*mot entier*	1e syll.	2e syll.
H					
M				———	
L	———	———			———

Tableau 22. Représentation schématique du passage de deux monosyllabes au ton L à un disyllabe dont les tons sont M et L, par l'intermédiaire d'un état (hypothétique) dans lequel le contraste tonal a lieu à l'échelle du mot.

4.2.2. Modifications de phonèmes

Outre les cas de modification tonale, des cas de dissimilation et d'assimilation phonémique sont sporadiquement observés. L'expression /mi ˊ dɯ̀/, « femme » + « grand », peut signifier « femme adulte, femme âgée », mais s'est également spécialisée dans le sens de « fille aînée » ; dans ce dernier sens, elle est parfois prononcée comme [mi ˊ dɯ̀ᵊ], avec diphtongaison centralisante du /ɯ/. La prononciation [mi ˊ dɯ̀ᵊ] ne peut signifier que « fille aînée », et non « femme adulte, femme âgée ». Le phénomène ne paraît pas encore lexicalisé. La modification des noms de mois (déjà évoquée ci-dessus, section 1.3.4) présente des similarités avec les phénomènes qui viennent d'être décrits : les trois premiers mois sont désignés par des noms qui ne sont pas décomposables de façon transparente (/jɤ̄ pè/, /hè ŋgjɤ̄/, /sá wā/) ; de même pour le dernier mois (12e mois), /ndā wā/ ; en

revanche, les noms de mois du 4e au 11e sont désignés par un nombre suivi de /mē/, et certains connaissent certaines modifications, comme le montre le tableau 23.

numéro	réalisation	modification par rapport à la simple juxtaposition du nombre et du mot /mē/
4e	lúᵊ mē	modification de la rime de « quatre »
5e	wɑ́ mē	pas de changement : « cinq » + « mois »
6e	ʈʂʰwɑ́ mē	pas de changement : « six » + « mois »
7e	ʂā mē	modification de la rime de « sept » : ɚ devient a
8e	hwɑ́ mē	modification de la rime de « huit » : o devient wɑ
9e	ŋgūᵊ mē	ŋgv̄ devient ŋgūᵊ
10e	tsʰè mē	pas de changement : « dix » + « mois »
11e	tsʰè djɤ̄	modification de la rime de « un » (onze se dit /tsʰē dù/, formé de /tsʰè/ « dix » et /dū/ « un », avec changement tonal, non /tsʰè dū/) : dū devient djɤ̄

Tableau 23. Les noms de mois dans le dialecte de AS, du quatrième au onzième mois.

La forme non modifiée n'est pas correcte : /lú mē/ signifie « quatre mois », non « le quatrième mois » ; /hó mē/ signifie « huit mois », non « le huitième mois ». Le changement a pour effet synchronique de distinguer nom de mois et nombre de mois[1].

A l'intérieur de certains mots lexicaux disyllabiques a lieu une modification de la rime de la première syllabe. Il peut s'agir d'une harmonie partielle : à la différence de /kv̩ fv̩/, « cheveux », formé de /kv̩/ « tête » + /fv̩/ « poil », le même /kv̩/ « tête » associé à /lȳ/ (classificateur des objets ronds) donne /kū lȳ/ « tête (partie du corps) ». Dans ce cas précis, le changement pourrait se décrire comme la transmission au /ɥ/ du trait d'arrondissement du /y/ de la seconde syllabe ; mais cette explication ne s'applique pas dans le cas de /kū ŋgɯ̄/, « oreiller », le /ɯ/ n'étant pas une voyelle arrondie.

L'harmonie peut également être complète, comme dans : /gū mū/ « corps », formé de /gɥ/ et /mu/ (Fang Guoyu et He Zhiwu 1995:420) ; et dans /bù hā/, « pâtée des cochons », formé de /bù/ « porc » et /hā/ « nourriture ».

Le fait que la rime /ɥ/ soit particulièrement sensible à la tendance à l'harmonie vocalique souligne son caractère relativement instable au plan phonétique.

Enfin, Fang Guoyu et He Zhiwu 1995:341 signalent un caractère signifiant « pierre-esprit », « esprit du seuil de la maison », qui se lit /dù lū/ ; la deuxième syllabe est probablement une

[1] Ce phénomène rappelle, de loin, la façon dont les noms de personnes sont formés en teko (langue tupi-guarani parlée en Guyane) : par une légère modification apportée à un nom d'animal. Ainsi, /kutʃili/ est un nom de personne formé à partir de /kutsili/, désignation d'une variété de singe ; le passage de /ts/ à /tʃ/ signale l'emploi détourné qui est fait du mot.

modification de /lɣ̩/, « pierre » ; il s'agirait alors d'une assimilation progressive d'arrondissement, modifiant /ɣ̩/ en /u/.

Pour résumer, les cas observés sont : le passage de /ɣ̩/ à /u/ avant *ou après* /u/, le passage de /ɣ̩/ à /u/ avant /ɯ/, et le passage de /u/ à /ɑ/ avant /ɑ/. L'inventaire demeure à compléter par la prise en compte d'un plus grand nombre d'exemples ; pour autant, il est peu probable que la force qui a modelé les mots en question soit la même dans tous les cas, et appelle description à l'aide de la seule notion d'assimilation. « An assimilatory process which may look phonetically plausible on the surface may in fact turn out to be something quite different » (Hyman 1975:249).

5. Variations intonatives dans la réalisation des phonèmes

L'inventaire des phonèmes d'une langue ne permet pas de prédire le détail phonétique de leur réalisation : la coordination des mouvements articulatoires est le résultat d'un long apprentissage, et est dans une certaine mesure spécifique à chaque langue (observation déjà formulée par Grammont 1933:22, *Traité de phonétique ;* voir, plus récemment, Keating 1990, Manuel 1999). Les observations qui suivent portent sur de tels détails phonétiques : co-articulation entre phonèmes, et variabilité en fonction du contexte prosodique. Ces observations non systématiques ont été rassemblées à mesure de l'étude de la langue, dans l'esprit d'une phonétique de la phrase (perspective qui rejoint également, nous semble-t-il, celle de Nolan 1999, « The devil is in the detail »).

Les mots grammaticaux connaissent une réalisation prosodiquement plus faible. Cette tendance, déjà relevée par Vaissière 1971 pour le français, et vérifiée dans de nombreuses langues, se traduit notamment par le **voisement des fricatives et affriquées non voisées en position non initiale**, et (de façon symétrique) par le **dévoisement des fricatives et affriquées voisées en position initiale**. Ainsi, la particule finale d'aspect accompli /sè/ est souvent réalisée [zè], le /s/ se voisant en position intervocalique. Ce voisement phonétique est indépendant du ton : le conditionnel/topicalisateur / ˊ sé/ est lui aussi fréquemment voisé en [zé]. Le dévoisement apparaît par exemple dans /zē tù/ « combien » en position initiale d'énoncé, par exemple dans l'énoncé suivant (tiré d'un dialogue traduit du chinois ; locuteur M4) :

(2)	zē tù	dū	k^hù	ŋi ¯	tsā wù lē ?
	combien	un	moment	nécessiter	*particules finales*
	« Combien de temps est-ce que ça prend ? »				

Le /zē/ initial est réalisé /sē/. Le dévoisement n'est pas réservé à la position initiale d'énoncé : il peut également marquer les frontières de groupes intonatifs, par exemple dans l'énoncé suivant :

(3) tʰɯ̄ zè´ ŋi¯ bɯ̄ ?

3e sg. lequel jour partir

« Quel jour part-il ? »

Le locuteur M4 a réalisé le /zè´/ comme [sè´]. Cette variante pourrait déboucher, en diachronie, sur un changement catégoriel.

L'interprétation paraît transparente : la réalisation voisée revient à ne pas interrompre le voisement, continuité qui contribue à marquer une continuité intonative ; la réalisation non voisée signale une frontière d'énoncé. Cette hypothèse reçoit confirmation d'exemples tirés de parole continue, qui montrent que la modification du trait de voisement de la consonne n'est pas l'effet mécanique d'une tendance phonétique, mais d'un procédé proprement *linguistique*. Ainsi, dans le récit « Coutumes » du locuteur M3 (27 secondes après le début du document) :

(4) kā nɤ̄ tʰɯ̄, ʤɻɯ̄ lò gɤ̄ gi¯ hò

autrefois DEICT./TOP. ville GEN. torrent, rivière

ʈʂʰɯ̄ ŋi¯ hò lò sé jɤ̄, (...)

DEICT./TOP deux[1] CLASSIFICATEUR dans COND./TOP. particule constative

« Autrefois, en ce qui concerne les canaux qui parcourent la ville (de Lijiang), ... »[2]

Une longue pause (1800 ms) précède le conditionnel/topicalisateur /sé/[3]. Celui-ci n'en est pas moins voisé de part en part, réalisé [zé.jɤ̄].

Il n'est pas exclu qu'une syllabe supplémentaire soit en fait présente : le conditionnel/topicalisateur /tʰɯ̄/ (ce qui donnerait : .../tʰɯ̄ sé jɤ̄/...), sous une forme extrêmement réduite ; les intuitions de locuteurs natifs ne sont pas unanimes à ce sujet. Que cette syllabe soit présente ou non, le voisement de la consonne /s/ ne peut être attribué de

---

[1] Ici, « deux » a le sens de « quelques » (au plan synchronique, il paraît bien s'agir d'un seul et même mot, mais il est imaginable que l'actuel /ŋī/ soit, au plan diachronique, le produit de la confusion de deux morphèmes : le chiffre « deux », et un morphème de pluriel/collectif /ni/ du type de celui qui existe en tamang, par exemple dans /ʔai-ni/ « vous autres », formé de /ʔai/, 2e pers. sg., et du pluriel/collectif /ni/ [Martine Mazaudon, communication personnelle]).

[2] La quantification /ʈʂʰɯ̄ ŋī hò/ (non indispensable au plan syntaxique) nous paraît véhiculer, en contexte, une nuance de familiarité : nous gloserions par « ces canaux bien de chez nous ». Dans la traduction, il a été choisi, par équivalence (terme entendu dans le sens où l'emploient Vinay et Darbelnet 1977, Chuquet et Paillard 1989), de rendre cette nuance par un verbe : « parcourent », plutôt que par « les quelques canaux », quantification qui en français nous paraît plus fortement liée à sa valeur littérale.

[3] Chez le locuteur M3, natif de AS mais qui a longuement séjourné dans diverses régions de la préfecture de Lijiang, et en particulier dans la ville de Lijiang, le conditionnel/topicalisateur ne porte pas de ton flottant. A Lijiang, ce morphème est réalisé /sé/ ; à AS, / ´ sé/.

façon mécanique à l'influence de la voyelle précédente, puisqu'au plan phonétique cette consonne n'est immédiatement précédée d'aucune voyelle. Le voisement constitue pour l'énonciateur une façon de rattacher la particule / ˊ sé/ à ce qui précède, comme si aucun intervalle de temps ne s'était écoulé. **Cette *figure de style phonétique*, passée dans l'usage courant en naxi, revient à employer le voisement comme signe de continuité**, valeur qui prend racine dans la tendance largement attestée que Thurneysen 1946 (cité par Martinet 1955:257-258) nomme « lénition », par laquelle toute consonne qui se trouve en position intervocalique voit son articulation affaiblie.

Par ailleurs, **en contexte intervocalique, la friction des fricatives et affriquées tend à s'atténuer**. L'enregistrement électroglottographique réalisé lors des séances d'élicitation de mots dans la phrase-cadre « Il écrit le caractère __ » (décrit en détail au chapitre III) donne l'occasion d'observer des réalisations très affaiblies de la séquence /ʈʂʰɯ̄.ndzʏ̀/ (déictique + « caractère »), dans lesquelles le voisement ne s'interrompt pas, l'affriquée aspirée /ʈʂʰ/ étant réalisée voisée, et sans occlusion (comme une approximante rétroflexe [ɻ̢]).

Sans multiplier les observations, dont la portée est parfois simplement anecdotique, signalons également la réalisation relativement fréquente de /i/ comme [e] dans les mots grammaticaux lorsque la syllabe qui suit comporte une voyelle autre que /i/ ou /y/. Le statut phonémique que la voyelle /e/ possède au même titre que /i/ n'empêche manifestement pas cette variation allophonique. Ainsi, la particule /pʰiˊ/, qui signale que le terme d'un procès est atteint et dépassé, est fréquemment associée au conditionnel/topicalisateur / ˊ sé/ ; l'une et l'autre connaissent une modification phonétique (qui au stade actuel ne paraît pas encore catégorielle) : réalisation [e] du /i/, qui représente une forme d'harmonie vocalique régressive entre les deux syllabes ; voisement de la consonne /s/ en /z/.

6. Aperçu de la variation dialectale : FK, NL

Le dialecte de AS est employé comme point de repère dans la description des phonèmes et tons des dialectes de FK et NL ; les brèves remarques ci-dessous se concentrent sur les différences constatées.

6.1. Dialecte de FK (locuteur M12)

6.1.1. Les initiales

6.1.1.1. Observations concernant les initiales coronales

Les initiales coronales sont très fortement rétroflexes devant /a/, /i/, /ɯ/, /u/, /wɑ/, /ɚ/ : elles sont réalisées [ʈʰ], [ʈ], [ɖ], [ɳɖ] ; le / l / est réalisé [ɭ], proche d'un *tap/flap* [ɾ]. La forte rétroflexion des /tʰ/, /t/, /d/, /nd/ a pour corrélat secondaire une certaine affrication.

Cette rétroflexion n'est pas présente devant les rimes /ɑ/, /o/, /y/, /ɣ/ et /e/.

Consonnes rétroflexes et non rétroflexes ne s'opposent que devant /ɤ/ ; l'opposition paraît peu productive. Quelques paires minimales existent : « bouillir (l'eau bout) » (fei^{4} 沸) se dit /tsɯ̀ tʰɤ́/ (avec pour variante /tsʰɯ̀ tʰɤ́/) ; /ʈʰɤ́/ signifie « couler goutte à goutte » (di^{1} 滴).

6.1.1.2. Observations concernant les consonnes bilabiales

Les initiales bilabiales suivies de /ɣ/ ne sont pas réalisées comme des *trills* : ainsi, /bɣ/ se réalise [bɣ], non [ʙɣ] comme c'est le cas à AS.

Les rimes /o/ et /u/ ne s'opposent pas après les bilabiales (tandis qu'elles s'opposent après les coronales et les vélaires) : seul /u/ apparaît.

6.1.2. Les rimes

6.1.2.1. Les rimes simples

La rime /ɚ/ est réalisée extrêmement fermée à FK. Après les fricatives et affriquées, elle se rapproche des allophones de /ɯ/ dans cette position. Inversement, les fricatives sommet de syllabe sont peu fricatives et se rapprochent d'approximantes : ces fricatives sont /ɣ/ et les allophones [ʐ̩] et [ʑ̩] de /ɯ/ après initiales fricatives et affriquées ; sons que Chao Yuen-ren note [ɻ] et [ʅ] en chinois, où ils sont allophones de /i/ dans cette même position. Le mot /ndzɯ̀/, « s'asseoir », a la même composition phonémique qu'à AS, NL, DYZ ; ne reconnaissant pas le degré de fermeture qui dans d'autres dialectes caractérise /ɯ/ après les fricatives et affriquées, j'ai d'abord été tenté de noter /ndzɤ̀/ à FK. De même, certains /lɣ/ ont été notés /lɤ/, de façon erronée. Au plan phonétique, **la distinction entre /lɣ/ et /lɤ/ est bien maintenue du fait que la combinaison /lɤ/ se réalise avec une forte rétroflexion de l'initiale** : [ɭɤ] ou [ɾɤ].

Le /a/ a tendance à être nasalisé après /h/ : /ha/ est réalisé /hã /.

Le /o/ est prononcé [o] (non [ø] comme à NL et DYZ), et ce en toutes positions, tandis qu'à AS cette voyelle est prononcée [o] sauf lorsqu'elle forme syllabe à elle seule ou suit un /h/, auxquels cas elle est prononcée [ø] (voir ci-dessus, section).

6.1.2.2. Les semi-voyelles

Les rimes complexes suivantes contiennent une semi-voyelle /w/ :

rime	exemple à FK	traduction	traduction en chinois	mot correspondant à AS
wɑ	ndwɑ́	étang, bassin	chi^{2} tang^{2} 池塘	ndwɤ̄
wɤ	kwɤ̀	se chauffer (auprès du feu)	kao^{3} 烤 (烤火)	(/ŋgjá/ ; n'est sans doute pas cognat)
we	wè dzé	oiseau	niao^{3} 鸟	ví zi^{-}
	ts^{h}wè tsʊ̄	*emprunt au chinois* marteau	chui^{2} zi 锤子	ʈʂ^{h}wè
we	ki ´ kwē	courge	gua^{1} 瓜	ki^{-} kwɤ̀
	gwè k^{h}i ´	chanson	min^{2} ge^{1} 民歌	(/ndzā pū/ ; n'est pas cognat)
we	wè wé	rond *(adjectif)*	yuan^{2} 圆 (球很圆)	wɤ́ wɤ̀

Tableau 24. Rimes comprenant une semi-voyelle /w/ dans le dialecte de FK.

La rime /we/ apparaît dans de nombreux emprunts au chinois. Dans certains mots, elle est une variante libre de la rime /wɤ/ : « raide, pentu » (dou^{3} 陡) peut se dire indifféremment /wɤ̀ bè$^{-}$/ ou /wē bè$^{-}$/. Il se pourrait que les /we/ actuels aient pour origine des /wɤ/, dont la prononciation aurait changé sous l'influence du nombre croissant d'emprunts chinois à rime /we/, mais il ne s'agit pour l'heure que d'une hypothèse.

S'agissant de la semi-voyelle /j/, la rime /jɤ/ est largement attestée. Le /y/ se diphtongue en [ju].

6.1.3. Les tons

6.1.3.1. Quatre tons lexicaux, plus un ton sur les emprunts chinois

Le dialecte de FK connaît trois tons lexicaux de fréquence comparable : H, M, L, et LM. Les emprunts chinois portent un ton distinct de ces quatre tons : un ton montant LH.

Le ton LM, fréquent dans le lexique, part d'une fréquence relativement basse, et monte peu. La durée des rimes portant ce ton ne paraît pas plus longue, à l'oreille, que celle des rimes portant un ton H, M ou L.

Le ton LH est typiquement plus long que les quatre autres (cette impression auditive n'a pas encore été vérifiée par le dépouillement des données électroglottographiques enregistrées à FK). Il passe du registre bas au registre haut.

6.1.3.2. Tons flottants

De même qu'à AS, la réduction de syllabes au ton haut conduit à la réassociation de leur ton à la syllabe précédente, ce qui, outre des contours LH, crée également des contours MH, séquence tonale distincte de la précédente. **Comme à AS, un test de perception a été réalisé,**

qui établit cette différence. La procédure a consisté à faire enregistrer par un même locuteur (M12) dix réalisations des deux énoncés suivants :

t^hɯ̄́ sé, ŋgù tsó ŋù. « Si (on) (le) boit, (on) aura mal (au ventre). »

(5a)	*t^hɯ̄*	*´ sé,*	*ŋgù*	*tsó*	*ŋù.*
	3e sg.	COND./TOP.	souffrir	TAM: futur	copule

« Elle/il va se faire mal. » (littéralement: « elle/il va souffrir »)

Le contexte fixé pour l'élicitation (en concertation avec l'informateur) était le suivant : un enfant fait quelque chose de dangereux, malgré les remarques qui lui ont été faites ; quelqu'un fait la réflexion : « Lui, il va finir par se faire mal ! »

(6a)	*t^hɯ̀*	*´sé,*	*ŋgù*	*tsó*	*ŋù.*
	boire	COND./TOP.	souffrir	TAM: futur	copule

« Si [agent non spécifié : toi, il...] boit, [toi, il...] va être malade. »

Le contexte fixé est celui d'une discussion au sujet d'un alcool frelaté.

La seule syllabe qui diffère d'un énoncé à l'autre est la première, au ton M dans un cas, au ton L dans l'autre.

Le test a consisté à présenter les stimuli deux fois en ordre aléatoire (avec des répétitions si l'auditeur le demandait), à l'aide d'un haut-parleur de bonne qualité. Dans ce test à choix forcé, l'auditeur devait choisir entre deux traductions en chinois : (5b) 喝的话，会疼的 (6b) 他呢，会疼的. Les réponses étaient notées par l'enquêteur. Le test a été mené deux fois, à trois jours d'intervalle, avec le locuteur qui avait enregistré les stimuli (M12) ; il a reconnu tous les stimuli. Un test d'identification a ensuite été mené avec deux auditeurs ; ceux-ci ont uniformément glosé par (6b) tous les stimuli, ne percevant pas de différence. Il semble que leur perception ait été orientée par la plausibilité sémantique plus grande de (6a), énoncé moins *marqué* au plan sémantique que (5a), lequel nécessite un contexte relativement spécifique. C'est seulement lorsque le test leur a été soumis à nouveau, selon la modalité du choix forcé, qu'ils ont identifié correctement tous les stimuli. Cette expérience établit que les contours LH et MH sont différents au plan de la perception aussi bien que de la production.

6.1.3.3. Variation expressive des tons

Les particules finales voient leur ton fréquemment modifié de façon non catégorielle par l'expressivité associée à leur contenu sémantique. Ainsi, le ton lexical de /mù¯/ et /mù¯/ est souvent difficile à reconnaître en contexte, ce qui confirme le statut prosodique particulier des particules finales.

Le ton H possède une variante expressive surélevée et fortement descendante, lexicalisée sur quelques items lexicaux à valeur expressive. David Bradley (séminaire « Langues tibéto-birmanes », LACITO-Sorbonne Nouvelle) rapporte que dans une langue ni (yi), le ton H du déictique peut être fortement rehaussé à des fins intensives, permettant de passer du sens de « là-bas » à « tout là-bas au loin », de « là-haut » à « tout là-haut là-haut ». En naxi de AS, le phénomène paraît lexicalisé sur un mot : /ló/ (« enjamber, franchir, traverser »), qui possède des emplois grammaticalisés (par exemple : /ló lɛ̄ bɯ̄/ « s'en retourner » ; /lɛ̄/ : « à nouveau » ; /bɯ̄/ : « aller »). Ce mot est nécessairement prononcé avec un contour très haut et fortement descendant, qui se distingue de celui des autres mots au ton H. En naxi de FK, ce sont trois adverbes qui présentent ce même ton *super-haut-descendant* lexicalisé : /dé/, pour quelque chose de lointain sans précision de direction (/dé ʈʂʰɯ̄ nɤ̀/ « là-bas au loin » ; /ʈʂʰɯ̄/ : déictique, /nɤ̀/ : « lieu, endroit ») ; /gɤ́/, pour quelque chose d'éloigné en contrehaut ; /mé/, pour quelque chose d'éloigné en contrebas. A l'opposé, aucun ton super-bas n'est apparu en naxi, ni (à notre connaissance) dans d'autres langues à tons, pour l'intensification d'un ton L. L'expérience menée au chapitre III confirme que l'intensification tend à rehausser tous les tons.

6.1.3.4. Phénomènes de sandhi

Dans le parler de FK comme dans celui de NL, le conditionnel/topicalisateur /ˊ sé/ se réalise comme un ton montant lorsqu'il est précédé d'une syllabe au ton haut. A NL, sa réalisation est alors [sèˊ] ; à FK, il se réalise comme [sè˗] dans ce contexte. La motivation de ce changement n'est pas encore éclaircie ; il pourrait s'agir d'une forme de dissimilation tonale ; mais les séquences de tons H sont courantes en naxi.

De même, les modifications de tons dans les séquences formées d'un nombre suivi d'un classificateur ne sont pas mécaniques : /sʊ̀/ « trois » associé au classificateur /mè/ (CL. des animaux) donne /sʊ́ mɛ̄/, mais /tsʰè/ « dix » associé à ce même classificateur donne simplement /tsʰè mè/. Les modifications tonales observées semblent tendre à l'évitement des schémas H + H, LM + H et H + M à l'échelle du mot composé, et à leur remplacement par L + H dans les deux premiers cas, par M + H dans le dernier. Dans le dialecte de NL, le schéma H + M est également rare : de nombreux mots ayant un schème tonal H + M dans les dialectes de DYZ et AS ont un schème M + H à NL. Les suites H + H qui résultent de l'association d'un nombre suivi d'un classificateur ont tendance à devenir H + M à AS, M + H à NL.

tons lexicaux des deux éléments associés	tendance à FK	tendance à NL	tendance à AS
H + H	L + H	M + H	H + M
LM + H		sans changement	sans changement
LM + L	L + L	sans changement	sans changement
M + H	M + LM	sans changement	sans changement
H + M	M + H	M + H	sans changement

Tableau 25. Principales tendances au changement tonal relevées dans les séquences nombre + classificateur.

Un des paradoxes de ce changement est que la modification de certaines séquences H + M en M + H à FK aboutit précisément à la création d'un schéma qui est par ailleurs modifié lorsqu'il résulte de l'association d'un mot au ton M et d'un mot au ton H. Cela ne représente pas un élément de circularité à proprement parler, mais suggère que ces modifications tonales ne sont pas l'effet de règles synchroniques qui auraient pour fonction de ramener les schémas tonals de mot à un inventaire restreint de gabarits privilégiés par la langue.

A la différence du dialecte de AS, à FK les tons ne sont pas modifiés lors de l'association d'un adjectif de couleur à un nom, sauf dans le cas de /ki ̀ sú/ « nuage » + /hɤ̃ ̀/ « rouge » qui donne /ki ́ hɤ̃ ̄/ ; la perte de sa deuxième syllabe est peut-être responsable du changement ; s'agissant d'un exemple isolé, il est difficile de trouver des éléments d'explication.

Il existe d'autres phénomènes marginaux de ce type, qui attendent pareillement explication. Ainsi, bien que les verbes au ton H se réduplicuent à l'identique, donnant une séquence H + H, le verbe /lỳ/, « voir », se réduplique en /lý lỳ/ « jeter un coup d'œil » ; le mot /hĩ ̄/ « homme » devient /hĩ ́/ en composition : l'autonyme des habitants de FK est /zɚ́ hĩ ́/ ; /kʰi ̄ kʰi ̄/ « vendre », /kʰi ̄ kʰi ̄ hĩ ́/ « colporteur, marchand » ; /hỳ/ « fer », /tɤ̄/ « frapper », /hỳ tɤ̄ hĩ ́/ « forgeron ».

6.2. Dialecte de NL (locuteur M10)

Le dialecte de NL est globalement proche des autres dialectes occidentaux observés (tels que DYZ et AS).

6.2.1. Les rimes

Le /o/ est réalisé [ø], comme à DYZ. A certains /o/ du dialecte de AS correspondent des /ɑ/ à NL : la différence portée par la rétroflexion de l'initiale à AS, opposant /lō/ « travail » et

/ɭō/ (2e syllabe du mot « os »), est, à NL, une opposition entre /lō/ (prononcé [lʊ̄]) et /lā/ : /ʂā lā/ « os », /lō bē/ « travailler ».

La voyelle /ɤ/ paraît occuper une partie de l'espace laissé vacant par le déplacement de /o/ vers [ø] : le /ɤ/ est prononcé plus proche de [o] que dans les autres dialectes étudiés.

6.2.2. Les initiales

Le système des initiales est semblable à celui du dialecte de DYZ. Prénasalisées /mb nd ŋg/ et voisées simples /b d g/ sont confondues en une seule série, et généralement prononcées avec une prénasalisation en forme de citation. Les initiales /l/ et /nd/ (cette dernière étant elle-même confondue avec /d/) sont confondues en /nd/.

6.2.3. Les tons

Les tons des monosyllabes sont dans l'ensemble les mêmes qu'à DYZ et AS. Sur les mots disyllabiques ou polysyllabiques, le schéma H + M est rare, bien qu'il ne soit pas totalement exclu, comme le montre l'exemple du mot /mɯ́ ndɚ̄/, « pousse de bambou » (zhu^{2} sun^{3} 竹笋). Le témoignage le plus productif de ce processus est le schéma de réduplication des monosyllabes à ton H, qui à AS donnent un composé réduplicé disyllabique de schéma tonal H + M, tandis qu'à NL le schéma est M + H. Le tableau 23 présente quelques exemples sur des mots lexicaux.

Mot à AS	Mot à NL	Traduction	Traduction chinoise
ki ˊ sā	ki ˉ sá	brouillard	wu^{4} 雾
sá wā	sā wá	le troisième mois [de même pour le 4e et le 8e]	san^{1} yue^{4} 三月
tsʰè ʈʂʰwá ŋi ˉ	tsʰè ʈʂʰwā ŋi ˊ	le seizième jour du mois	shi^{2} liu^{4} ri^{4} 十六日
ɣ́ zi ˉ	ɣ̄ zi ˊ	oiseau	niao^{3} 鸟
ki ˊ ʂɤ̄	ki ˉ ʂɤ́	merle	xi^{3} que^{4} 喜雀

Tableau 26. Exemples de correspondances entre un schéma tonal de disyllabe H + M à AS et M + H à NL.

6.3. Eléments de comparaison entre dialectes

En vue d'une future étude des correspondances entre dialectes naxi, et avec les langues voisines (en particulier les langues yi-birmanes, sur la base desquelles des reconstructions ont été proposées par Burling 1967 et Bradley 1979), voici quelques premières remarques sur les dialectes naxi.

6.3.1. Généralités

Les initiales sont relativement stables d'un dialecte à l'autre, mis à part l'affaiblissement de certains mots grammaticaux : ainsi, à l'interrogatif /dè/ de FK correspond /zè/ à AS (ainsi que /zē/ et /zè˝/, dans /zē tù/ « combien », /zè˝ kv́/ « où »), passage d'occlusive à fricative.

Outre les faits rapportés en section 1.2 concernant les initiales rétroflexes, certaines correspondances entre syllabes à initiale rétroflexe suivie de /ɣ/ à AS et initiale non rétroflexe suivie de /ɯ/ à DYZ devront faire l'objet d'analyses comparatives prenant en compte plusieurs dialectes.

mot	**AS**	**Yang[4] Xi[1] 漾西**	**NL**	**FK**
hiver (冬)	mɯ̄ ts^{h}ɯ̄	mɯ̄ ts^{h}ɯ̄	mɯ̄ ts^{h}ɯ̄	mú ts^{h}ú
automne (秋)	mɯ̄ ʈʂ^{h}ý	mɯ̄ ʈʂ^{h}ý	mɯ̄ ts^{h}ú	(non utilisé)
pêche (桃子)	bɣ̄ ɖʐɣ̄	bɣ̄ dzɣ̄	bɣ̄ dzɯ̄	bý ki˘ ¯
sueur (汗)	ʈʂý	tsý	tsú	ʈʂù ¯
faucille (镰刀)	ʂý kɣ̄	ʂý kɣ̄	sɯ̄ ký	ʂù gý
clef (钥匙)	ɳɖʐɣ̄ kɣ̀	ndzɯ̄ kɣ̀	ndzɯ̄ kɣ̀	(tsè kə̄˗)
chemin (路)	ʐɣ̄ gɣ̄	zɯ̄ gɣ̄	zɯ̄ gɣ̄	ʐɤ̀ gý
bilan : motif le plus courant	rétroflexe + /ɣ/	rétroflexe ou non rétroflexe + /ɣ/	non rétroflexe + /ɯ/	rétroflexe + /ɯ/

Tableau 27. Illustration des deux correspondances à AS des syllabes à initiale non rétroflexe suivie de /ɯ/ d'autres dialectes : initiale rétroflexe suivie de /ɣ/, initiale non rétroflexe suivie de /ɯ/.

L'absence du terme « automne » à FK s'explique par le fait que l'année est traditionnellement divisée en deux saisons, non en quatre : l'été (/mú ʐú/), saison des pluies, et l'hiver (/mú ts^{h}ú/), saison sèche.

Les correspondances de voyelles sont très variées ; le premier survol proposé dans le tableau 28 en donne une idée, sans entrer dans le détail des contextes conditionnants.

Rime à FK	Rime à AS
u	ɣ (305, 339, 788), u (193, 204), ɯ (1)
i	ɯ (22, 42, 50, 187, 362, 674, 1016), i (2), ɣ (390)
e	i (216, 646, 826, 959), e (2, 166, 214), a (1160)
a	ə˞ (7 hə̄˞ vent, 30, 194, 213, 338, 732, 745), a (20, 136), ɯ (13), e (2, 403), o (237), ɑ (402), ja (404, 424)
ɑ	ɑ (124, 200), e (337 ; cognat incertain)
o	o (260, 357), ɣ (43, 177)
ə˞	o (16, 27, 158, 196, 270, 325, 335, 371, 463, 893), ə˞ (220, 311, 931), a (160, 243, 1016), ɯ (107, 720), ɣ (156), u (710)
ɯ	ə˞ (81, 648, 655), ɣ (247, 339, 367), ɯ (195), a (13, 423), u (1, 8, 62), y (659)
v	ɣ (30, 174), y (181, 733, 826), ɯ (276, 326, 656, 656), ə˞ (333, 413), o (493)
ɤ	ɣ (40, 427, 844, 852), ɤ (36, 1037)
y	y (52, 108), ɣ (230), i (696)
jɤ	jɤ (66, 191), ɯ (795)
ja	e (315)
wɑ	wɑ (103, 349), a (110), u (119, 783), o (177)

Tableau 28. Principales correspondances entre rimes des dialectes de FK et AS. Les numéros entre parenthèses renvoient aux items de notre liste d'environ 1500 mots (collectés de première main).

Les correspondances tonales n'ont pas encore été établies systématiquement. Pour l'essentiel, mis à part l'inversion en M + H, à NL, des séquences H + M sur disyllabes observées à AS et DYZ, les tons sont les mêmes à AS et NL. Le dialecte de FK présente en revanche des correspondances beaucoup plus complexes. Un premier aperçu suggère qu'un grand nombre de syllabes au ton M de AS correspondent à des syllabes à ton H à FK, et un grand nombre de H de AS correspondent à des M à FK ; beaucoup de disyllabes M + L de AS correspondent à des disyllabes H + M à FK.

Une reconstruction systématique du naxi reste à faire. Un tableau des combinaisons attestées entre initiales et rimes a été préparé, mais n'est pas présenté ici, car les questions qu'il soulève ne peuvent être traitées que dans le cadre d'un travail de comparaison approfondi : ainsi, la rime /o/ n'apparaît pas après les bilabiales, mis à part certains emprunts au chinois (Fang Guoyu et He Zhiwu 1995:371, 384) et deux exemples à vérifier, l'un fourni par Fang Guoyu et He Zhiwu 1995 : /pʰó/, traduit par « montrer, conseiller, instruire » (指点), l'autre par Pinson 1998 : /pʰó sū/ « grosse grenouille, grenouille-buffle » (牛蛙) ; l'interprétation de ce fait demande que les sources du /o/ du dialecte de AS aient préalablement été précisées.

6.3.2. Distinctions perdues et confusions en cours dans le dialecte de DYZ et autour de la ville de Lijiang

6.3.2.1. Affriquées et fricatives rétroflexes et non rétroflexes

Le consultant M2 (He Jiren) pense constater l'extension, dans son dialecte (Yang[4] Xi[1] 漾西), des affriquées et fricatives non rétroflexes aux dépens des rétroflexes. La distinction serait ainsi en passe de disparaître par diminution du nombre d'items prononcés avec rétroflexion.

La locutrice F1, originaire de AS mais qui habite depuis plusieurs années à Kunming, capitale de la province, et habite dans la ville de Lijiang plutôt qu'au village de AS lorsqu'elle revient au pays, a conscience de perdre peu à peu l'opposition entre rétroflexes et non rétroflexes. Elle prononce « dire » /sʏ́/, tout en se souvenant que la forme employée à AS est /ʂʏ́/ ; dans d'autres cas, elle prête au dialecte de AS des rétroflexes sur des mots qui en réalité n'en présentent pas : ainsi, elle explique que « dix » se dit /ts^{h}è/ à la ville (DYZ), /ʈʂ^{h}è/ au village de AS ; en réalité, ce mot se dit /ts^{h}è/ dans l'une et l'autre localité. Cette hyper-correction confirme la perte de l'opposition dans son parler.

Dans les données de Mazaudon et Michailovsky 1979, l'opposition est présente, mais un grand nombre d'items ayant une initiale rétroflexe dans nos données n'en ont pas dans le parler décrit par Mazaudon et Michailovsky 1979 : /swɑ̀/ « grand », contre /ʂwɑ̀/ ; /sù/ « fer » (homophone : « chercher »), contre /ʂù/. Dans ces données, l'affaiblissement de l'opposition est tel que rétroflexes et non rétroflexes paraissent être en distribution complémentaire : rétroflexes [ɖʐ], [ʈʂ] et [ʈʂ^{h}] devant /ɯ/ et /ɚ/ ; [dz], [ts] et [ts^{h}] devant toutes les autres voyelles. (Dans nos données, ces deux séries s'opposent : par exemple, /ʈʂ^{h}ɯ̄/ est le déictique proche, /ts^{h}ɯ̄ tè/ « couteau » ; /sò ŋi˧/ « demain », /ʂō ʅō/ « os » ; de même pour l'ensemble / ʈʂ^{h} ʈʂ ɖʐ ɳɖʐ ʂ ʐ/ s'opposant à /ts^{h} ts dz ndz s z/.)

La préservation de certaines oppositions à DYZ rappelle, en français, les exemples souvent cités des paires de mot « patte » /pat/ et « pâte » /pɑːt/, « maître » /mɛːtʁ/ et « mettre » /mɛtʁ/ : les deux membres de chacune de ces paires seraient confondus par la plupart des locuteurs du français parisien cependant que certains locuteurs plus âgés conserveraient l'opposition (qui se réalise au plan phonétique par une différence de qualité vocalique pour a-ɑ, et de durée pour l'une et l'autre opposition). En dehors des quelques paires minimales souvent citées, les locuteurs qui pensent maintenir les distinctions et déplorent leur perte dans la jeune génération les ont en fait généralement perdues, et prononcent /a/ la voyelle de « laque » comme celle de « lac » (voir également Martinet 1945, Martinet et Walter 1973).

L'application mécanique du principe de reconstruction amènerait à postuler des origines séparées pour l'initiale selon qu'elle présente une correspondance (entre DYZ et AS) non rétroflexe – rétroflexe, non rétroflexe – non rétroflexe ou rétroflexe – rétroflexe. Dans l'état

actuel des connaissances, il n'est pas possible de trancher entre cette hypothèse (qui ne va pas de soi) et celle (qui nous paraît plus vraisemblable, mais reste à étayer par des observations affinées) d'une perte *partielle* des oppositions entre rétroflexes et non rétroflexes dans le dialecte de DYZ, dans un contexte de contact entre dialectes.

6.3.2.2. Occlusives et affriquées prénasalisées et voisées

L'opposition entre occlusives et affriquées prénasalisées et voisées est entièrement absente du dialecte de NL, où le produit de la confusion est généralement prononcé prénasalisé. Thomas Pinson a fait le même constat dans les dialectes de la plaine de Lijiang qu'il a étudiés (Pinson 1996, 1998). He Jiren et Jiang Zhuyi 1985 décrivent l'opposition comme perdue dans le parler de la ville (DYZ), et ne l'incluent pas dans la notation romanisée du naxi : ainsi, les /gūū/ et /ŋgūū/ de AS se transcrivent pareillement /kee/ (ce qui a amené He Xueguang, dont le dialecte, AS, conserve cette distinction, à l'ajouter dans sa version modifiée de la romanisation naxi).

Certains locuteurs habitant à Lijiang conservent néanmoins l'opposition (par exemple M6) ; le dialecte de la ville n'est pas homogène. Certains locuteurs ont perdu l'opposition lors d'un changement d'environnement linguistique (par exemple la fille de M4, née à AS, et qui habite NL depuis une quinzaine d'années), d'autres généralisent peu à peu la prononciation prénasalisée, tout en conservant la distinction sur certains mots : ainsi, F1, née à AS et qui habite DYZ (et la capitale de la province, Kunming) depuis plusieurs années, prononce /m͡bù/ « cochon » (AS : /bù/) ; un des consultants de Mazaudon et Michailovsky 1979, qui conserve par ailleurs assez bien la distinction, prononce /m͡bə̄/ « invité, hôte » (AS : /bə̄/).

6.4. Vérifications à effectuer dans le dialecte de Zhong¹He² 中和, décrit par Fu Maoji 1981

Fu Maoji 1981 distingue, outre /ɑ/ et /a/, un /æ/, avec la paire minimale suivante : /hæ/ « chaux » (石灰), /hà/ « or » (金子) (homophone « acheter », 买). Dans les dialectes étudiés de première main, et dans les diverses publications, ces mots sont homophones. De l'avis du linguiste He Jiren, il s'agirait d'une erreur de Fu Maoji, mais le fait reste à vérifier dans le dialecte concerné (zhong¹ he² 中和).

Le même ouvrage oppose en outre une initiale /ɣ/ à l'absence d'initiale. Tous les exemples à initiale /ɣ/ fournis correspondent (dans nos données, et dans celles des autres auteurs) à des syllabes sans initiale : il s'agit (dans notre notation) de : /ūū/ « vache », /ù/ « esclave » (Fang Guoyu et He Zhiwu 1995:211), /ò/ « oie », /à/ « poule », /ù/ « se réunir, se rassembler, se rencontrer ». Les exemples sans initiale donnés par Fu Maoji 1981 sont : /ū/, « nom d'un bouddha », /ò/ « nom d'esprit » (mots dont nous n'avons pas trouvé d'équivalent), et /á/ « oie », qui dans nos données est également /á/, même syllabe (au ton près) que le /ù/ « se rassembler » que Fu Maoji note /ɣù/. L'existence de ces différences pourrait être un artefact

dû à l'expérience linguistique complexe de l'informateur de Fu Maoji (déjà évoquée ci-dessus, section 1.2.3).

Eléments de conclusion

L'entreprise de reconstruction d'états anciens du naxi n'en est qu'à ses débuts. Aucun des dialectes naxi étudiés à ce jour ne se révèle spectaculairement conservateur ; chacun des dialectes n'en offre pas moins des indices pour la reconstruction. Le dialecte de FK, qui de par sa situation géographique périphérique pourrait apparaître comme conservateur, a en réalité perdu des oppositions qu'il y a de bonnes raisons de croire anciennes ; il possède quatre tons lexicaux de fréquence comparable, ce qui représente une innovation par rapport au système tonal des dialectes de AS, NL et DYZ. Le dialecte de AS apparaît relativement conservateur ; celui de NL, moins conservateur, et proche du parler de la ville de Lijiang (dialecte de DYZ), présente des correspondances atypiques de consonnes avec les autres dialectes qui pourront être utiles à la comparaison, même si leur analyse promet d'être complexe.

Pour la poursuite du travail, une extension de l'enquête aux parlers naxi orientaux (aussi appelés mo^{2} suo^{1} 摩梭语) est envisagée.

Table des matières de l'Annexe 2 (« Résultats expérimentaux complémentaires, et outils informatiques »)

1. Données anglaises complémentaires ____ 381

1.1. Résultats des mesures sur les données de l'expérience-pilote avec le locuteur anglais M1 ____ 381

1.2. Réalisation des consonnes initiales *lenis* et *fortis* ____ 383

1.3. Nécessité d'exclure les consonnes continues voisées initiales de syllabe ____ 383

1.3.1. Problématique générale : la division de la syllabe en initiale et rime ____ 384

1.3.2. Arguments fournis par les données anglaises ____ 385

1.4. Caractéristiques intrinsèques des phonèmes vocaliques ____ 389

2. Résultats de mesures de débit d'air des tons vietnamiens B2, D1 et D2 ____ 391

2.1. Déroulement de l'enregistrement ____ 392

2.2. Pose des bornes de début et fin ____ 396

2.3. Résultats de débit d'air oral ____ 396

3. Etude de DECPA, amplitude des pics de fermeture sur le signal électroglottographique ____ 399

3.1. Introduction : la question de la mesure d'intensité, et l'utilité de mesures de la pente spectrale ____ 400

3.2. A quoi correspond le paramètre DECPA ? ____ 402

4. Scripts développés dans l'environnement MATLAB ____ 404

4.1. Outil d'analyse du signal électroglottographique : implémentation logicielle de la méthode de détection des pics positifs et négatifs sur la dérivée du signal ____ 405

4.1.1. Programmes existants au début de notre recherche ____ 405

4.1.1.1. Programmes de Nathalie Henrich ____ 405

4.1.1.2. Programmes de Vu-Ngoc Tuân ____ 406

4.1.2. Orientations du développement logiciel effectué au cours de la présente étude ____ 408

4.1.2.1. Choix d'une méthode semi-automatique ____ 408

4.1.2.2. Choix de ne pas appliquer de traitement sur le signal enregistré ____ 408

4.1.3. Principes pour la détection des pics de fermeture ____ 409

Traitement des pics de fermeture doubles ou multiples ____ 410

4.1.4. Principes pour la détection des pics d'ouverture ____ 410

1.4.1. Choix de lisser le signal ____ 411

1.4.2. Détection du minimum local ou barycentre des pics : Choix visuel au vu de quatre résultats ____ 412

4.1.5. Réflexion sur les méthodes de détection des pics de fermeture pour le calcul de la fréquence fondamentale. Proposition de deux algorithmes. ____ 415

Conclusion au sujet de la fréquence fondamentale ____ 417

4.2. Interface de vérification de valeurs de formants ____ 418

4.3. Implémentation d'un test-t apparié ____ 418

4.4. Simulations syllabiques ____ 419

4.4.1. Explication au sujet du nombre de points des courbes rééchantillonnées ____ 419

4.4.2. Traitement des cas où certains points de données manquent ____ 419

4.4.3. Des moyennes aux simulations syllabiques ____ 420

4.4.4. Perspectives de développement futur ____ 421

Annexe 2. Résultats expérimentaux complémentaires, et outils informatiques

1. Données anglaises complémentaires

1.1. Résultats des mesures sur les données de l'expérience-pilote avec le locuteur anglais M1

L'expérience-pilote menée avec le locuteur anglais M1 n'est pas rapportée dans le corps de la thèse, du fait que le corpus employé n'est pas le même que pour les étapes ultérieures de l'enquête, décrites au chapitre III de la thèse. Les résultats de cette expérience n'en présentent pas moins un intérêt : ils nous paraissent appuyer les conclusions tirées des données des locuteurs M3 à M7.

Les résultats des mesures sur les données de l'expérience-pilote avec le locuteur anglais M1 sont présentés dans les tableaux 26 à 30, et sur la figure 2.1. Celle-ci représente fréquence fondamental et quotient ouvert en fonction du temps.

syllabe	F_0 moyen (Hz) et écart-type, cond. C	F_0 moyen (Hz) et écart-type, cond. E	différence : E moins C, en % (gauche) et valeur approchante en tons musicaux (droite)		partie de la syllabe (sur les 100 échantillons également espacés) sur laquelle la différence est significative	diff. d'écart-type de F_0: E comparé à C, en %
accentuée	110 (7.2)	140 (12.4)	27	2 ¼	1 à 100	+72
non accentée	89 (4.5)	91 (6.0)	(=)	(=)	1 à 40, 72 à 100 (Fo plus haut en condition E)	+33

Tableau 26. Valeurs de fréquence fondamentale des disyllabes du locuteur M1 en conditions C et E.

La différence globale de F_0 sur la syllabe non accentuée est très faible. La fréquence fondamentale est néanmoins significativement supérieure sur les premiers 2/5 de la syllabe. Le contraste de hauteur entre les deux syllabes est accru en condition E, du fait que la première syllabe est nettement plus haute et la seconde peu modifiée.

syllabe	évolution de F_0 du 5e au 95e point (en tons musicaux)		statistiques sur la différence d'évolution de F_0 entre C et E	différence d'évolution (dans les cas significatifs): condition E comparée à C
	condition C	condition E		
accentuée	-1.3	-1.0	ns ($t_{78} = 1.24$, $p = 0.21$)	
non accentuée	-0.7	-1.0	s ($t_{78} = -3.2$, $p = 0.002$)	descente plus forte

Tableau 27. Evolution de la fréquence fondamentale des disyllabes du locuteur M1 en conditions C et E.

syllabe	quotient ouvert moyen (%), cond. C	quotient ouvert moyen (%), cond. E	différence entre E et C (%)	partie de la syllabe (sur les 100 échantillons également espacés) sur laquelle la différence est significative	écart-type de E comparé à C (%)	différence (dans les cas significatifs): condition E comparée à C
accentuée	57.4	53.4	-7	1 à 100	(=)	plus bas
non accentuée	63.6	54.1	-18	1 à 100	-43	plus bas

Tableau 28. Quotient ouvert des disyllabes du locuteur M1 en conditions C et E.

Au bilan, **la différence de quotient ouvert est significative sur les deux syllabes.** Elle est particulièrement nette sur la deuxième syllabe : la qualité de voix de la seconde syllabe est détendue en condition C, tandis qu'elle est très proche de celle de la première syllabe en condition E. **Cette observation peut s'interpréter comme une atténuation, en condition E, de la réduction qui frappe la syllabe non accentuée.**

syllabe	longueur moyenne (ms), cond. C	longueur moyenne (ms), cond. E	différence: E moins C, en %	résultats du test-t sur la longueur (valeur de t_{78} et p)	différence (dans les cas significatifs): condition E comparée à C
accentuée	145	155	+7	ns ($t_{78} = -1.9$, $p = 0.058$)	
non accentuée	152	215	+41	s ($t_{78} = -9.86$, $p < 10^{-30}$)	plus long

Tableau 29. Longueur des disyllabes du locuteur M1 en conditions C et E.

syllabe	F_0 moyen: partie de la syllabe sur laquelle la diff. est significative	amplitude de la diminution de F_0	longueur syllabique	quotient ouvert: partie de la syllabe sur laquelle la diff. est significative
accentuée	totalité	ns	ns	totalité
non accentuée	2/3	s	s	totalité

Tableau 30. Récapitulatif des différences entre conditions pour les disyllabes du locuteur M1.

Dans le tableau 30, le seuil suivant a été choisi (empiriquement) concernant l'indication globale d'un F_0 moyen différent sur la syllabe : la différence de F_0 moyen est indiquée « s »

si elle est significative sur plus de 75% de la syllabe. Ce choix se justifie par le fait que cette mesure vise à refléter le registre global, non les mouvements mélodiques.

1.2. Réalisation des consonnes initiales *lenis* et *fortis*

Ce paragraphe revient brièvement sur la réalisation phonétique des consonnes anglaises communément notées comme /b/, /d/, /g/, /p/, /t/, /k/.

Le placement des balises de début et de fin pour l'analyse du signal électroglottographique a fourni l'occasion d'observations sur la réalisation des consonnes initiales transcrites comme /b d g p t k/ (notation traditionnelle qui est retenue dans la présente étude dans le souci de prévenir des confusions). **Dans les deux conditions de lecture, en position initiale d'un mot accentué (monosyllabique ou disyllabique), aucune de ces consonnes n'est voisée : le voisement coïncide avec le début de la voyelle.** Dans les termes proposés par l'Alphabet Phonétique International (International Phonetic Association 1999), ces observations peuvent se résumer de la façon suivante : l'une des séries de consonnes est réalisée *non voisée*, l'autre est réalisée *non voisée aspirée*, comme l'indique le tableau 31.

catégorie phonologique (transcription usuelle)	b	d	g	p	t	k
transcription phonétique la plus adéquate pour les données enregistrées	p	t	k	p^h	t^h	k^h

Tableau 31. Réalisation des occlusives initiales dans les données anglaises recueillies.

Ce fait est bien connu des spécialistes : voir par exemple l'article « Voiced and voiceless » du dictionnaire de prononciation de John Wells (Wells 1990:767), où il est suggéré que l'opposition des deux séries en anglais est à concevoir en termes de *lenis* vs. *fortis*, plutôt qu'en termes d'une stricte opposition de voisement[1]. Dans la perspective d'une comparaison des résultats entre langues, il importait de rappeler ce fait que la transcription usuelle peut avoir tendance à faire oublier.

1.3. Nécessité d'exclure les consonnes continues voisées initiales de syllabe

Les données anglaises fournissent un argument en faveur de l'idée traditionnelle selon laquelle la partie significative de la courbe de hauteur est portée par la rime, à l'exclusion des consonnes initiales.

[1] Au sujet des catégories *lenis* et *fortis*, qui remontent aux travaux de linguistes du XIXe siècle, voir notamment Kohler 1979 « Dimensions in the Perception of Fortis and Lenis Plosives », et Kohler 1984 « Phonetic explanation in phonology: the feature fortis/lenis ».

1.3.1. Problématique générale : la division de la syllabe en initiale et rime

Dans l'étude des langues d'Asie du Sud-Est, la division en initiale et rime est ancienne ; le ton appartient à la syllabe, mais est porté par la rime. Cette division ressort de l'analyse pratiquée de longue date dans les dictionnaires de rimes de la tradition linguistique chinoise[1]. Dans une syllabe ayant pour phonèmes /t^hu/, l'initiale est /t^h/ et la rime /u/ ; dans /t^han/, l'initiale est /t^h/ et la rime /an/ ; dans /man/, l'initiale est /m/ et la rime /an/. **Le ton est porté par la rime. Dans le vocabulaire actuel** (par exemple : Goldsmith 1976, Clements et Ford 1979:181), **on dira que l'unité porteuse de ton *(Tone-Bearing Unit)* est la rime.** Dans la perception qu'en ont les locuteurs des langues modernes, par exemple le chinois mandarin et le vietnamien, il est malaisé de distinguer la part qui revient à la structure linguistique proprement dite et celle de l'apprentissage : les écoliers de Chine ou du Vietnam apprennent en effet un équivalent du B-A BA auquel s'ajoute le ton. La division en initiale, rime et ton est la base de l'apprentissage scolaire de la norme de prononciation, et de l'écriture romanisée : alphabet *pinyin* en Chine continentale, et alphabet *quốc ngữ* au Vietnam, où il est l'orthographe nationale depuis la première moitié du vingtième siècle. Les descriptions de nombreuses langues confirment la division en initiale et rime (par exemple dans les langues yi/ni : « Syllables with final nasals behave tonemically like open ones » : Burling 1967:56). Dans le domaine africain, pour ne citer qu'un exemple, Yetunde Laniran mesure la partie vocalique de chaque syllabe yorùbá :

> Measurements were taken from the vowel portion alone for a number of reasons : 1) In cases where there is a voiceless consonant, the F_0 value cannot be measured ; 2) There is no evidence in Yorùbá to suggest that targets are required in consonants. The F_0 value within the consonant can be generated by interpolation, once target values for the F_0 in adjacent vowels are specified. (Laniran 1992:61)

Cette division n'est pas spécifique aux langues à tons. Elle est également rapportée par Gårding 1998:137 pour le danois[2]. Pike 1948:30 signale n'avoir jamais rencontré de phénomène mélodique significatif, ni tonal ni intonatif, porté par une occlusive ou fricative

---

[1] Les ouvrages chinois sur le sujet sont très nombreux ; citons, parmi les travaux qui mettent en regard tradition linguistique chinoise et linguistique (occidentale) moderne, le recueil d'articles de Xue Fengsheng (1999) : 汉语音韵史十讲，薛凤生著，耿振生　杨亦鸣选编，华语教育出版社，北京, 1999.
Si l'on détaille, la syllabe est divisée en quatre parties : l'initiale, 声母 ; la semi-voyelle éventuelle, appelée (si l'on traduit littéralement) « tête de la rime », 韵头 ; le noyau vocalique, littéralement « tronc/ventre de la rime », 韵腹 ; et la consonne ou semi-voyelle finale, « queue de la rime », 韵尾.

[2] Gårding attribue l'observation à Selmer 1928 ; l'article étant en danois, nous n'avons pu en prendre connaissance.

voisée initiale (en revanche, il ne se prononce pas sur le cas des continues telles que les nasales).

Cette idée a récemment été remise en question par une floraison d'études réalisées dans le cadre autosegmental-métrique : Xu Yi et Wang 2001:321, Xu Yi 1998 pour le chinois, Ladd, Faulkner *et al.* 1999 pour l'anglais, Arvaniti, Ladd *et al.* 1998 pour le grec, Prieto, van Santen *et al.* 1995 pour l'espagnol mexicain. **Ces auteurs jugent que la courbe mélodique portée par les consonnes initiales voisées continues (typiquement /l/, /m/, /n/, /ɲ/) est partie intégrante de la courbe de fréquence fondamentale qui caractérise la syllabe dans son entier.** Ainsi, une syllabe [ma] dont la fréquence fondamentale s'élève au cours du [m] et descend au cours du [a] est considérée comme portant un schéma mélodique montant-descendant, tandis que dans le cadre d'une division entre initiale d'une part, rime porteuse du ton d'autre part, seule compte la descente sur la rime [a]. De nombreux auteurs (dont Ladd et Pierrehumbert) recourent à des consonnes initiales continues, /m/, /n/, /ɲ/, /l/, qui à la différence des occlusives n'ont pas d'effet visible sur la courbe de fréquence fondamentale des voyelles, et permettent d'obtenir des courbes de fréquence fondamentale continues sur la syllabe.

Xu Yi va jusqu'à défendre l'idée qu'en chinois mandarin, l'alignement aurait lieu par rapport au début de la syllabe, le ton étant porté par la syllabe tout entière : « the syllable is presumably the host for tonal targets » (Xu Yi et Wang 2001:321 ; voir également Xu Yi 1998), ce qui est en contradiction avec les observations, sur cette même langue, de Howie 1974 (« On the domain of tone in Mandarin »), Sagart 1993 (dans les dialectes gan) et Hallé 1994[1]. (Au sujet de la notion d'*alignement de la courbe de fréquence fondamentale avec l'articulation*, voir l'Annexe 3, section 1.)

1.3.2. Arguments fournis par les données anglaises

L'expérience proposée ici pour apporter un argument en faveur de la division de la syllabe en initiale et rime s'appuie sur les données anglaises : la composition du corpus anglais permet de comparer les courbes de trois catégories de syllabes, celles comportant une initiale occlusive *lenis* /b/, /d/ ou /g/ (phonétiquement non voisée dans ce contexte), celles comportant une initiale occlusive *fortis* /p/, /k/ ou /k/ (phonétiquement aspirée), et celles comportant une continue sonore initiale /m/ ou /n/ (en anglais, /ŋ/ n'apparaît pas en position initiale de syllabe). La figure 2.2 montre, pour un locuteur (M4) et dans une même condition

[1] Une présentation de John Ohala et de David Mortensen dans le cadre des vidéo-conférences Berkeley-Paris argumentait pareillement en faveur d'une division en initiale et rime (cette dernière incluant les consonnes nasales finales) à partir de données d'une langue de la branche hmông de la famille hmông-miên.

de lecture (condition S), les courbes brutes de ces trois types de monosyllabes. L'inspection visuelle des courbes révèle la relative homogénéité de chacun de ces trois ensembles.

La comparaison des courbes des labiales, coronales et vélaires montre une grande ressemblance des courbes de F_0 ; il n'a donc pas paru nécessaire de signaler le lieu d'articulation sur la figure. Le comportement des vélaires est proche de celui des labiales et coronales ; /k/ et /g/ sont donc inclus dans les figures et les calculs statistiques malgré l'absence de nasale /ŋ/ correspondante.

Les courbes des syllabes comportant une syllabe initiale /p/, /t/, /k/ (en bleu) ont une allure globalement descendante. (La valeur très élevée à la fin de trois des courbes est due à une période de plus petite amplitude qui se trouve juste avant l'interruption totale du voisement.) Les courbes des syllabes comportant une consonne initiale /b/, /d/, /g/ (en noir) comportent une courte montée en début de voisement (d'environ 25 ms). Les courbes réalisées sur la longueur totale des syllabes incluant une initiale voisée continue présentent une allure différente de celle des deux ensembles précédents : une longue montée précède la descente[1].

Les figures 2.3 à 2.10 mettent en regard, pour chaque locuteur et chacune des deux conditions de lecture, quatre courbes : la courbe moyenne obtenue sur les syllabes à initiale *lenis* /b/ /d/ /g/ ; celle correspondant aux initiales *fortis* /p/, /t/, /k/ ; celle obtenue sur les syllabes à initiale continue voisée (/m/ ou /n/), en incluant la totalité du voisement (correspondant à la consonne initiale plus la voyelle) ; et enfin la courbe obtenue sur ces mêmes syllabes à initiale continue voisée lorsque seule la partie vocalique est prise en compte. Afin d'évaluer le degré de proximité entre les courbes sur les deux types de rimes non précédées d'une initiale continue voisée et le type de rime obtenu en ne prenant pas en compte la consonne initiale continue voisée, ces trois types de rimes ont été alignés par leur point d'origine.

Seules les courbes des monosyllabes sont présentées ; les résultats pour les disyllabes sont extrêmement similaires.

---

[1] Sur la figure, trois des points terminaux des courbes de fréquence fondamentale se situent dans une plage très élevée, valeur double voire quintuple de celle qui précède. Il ne s'agit pas d'une erreur de mesure, mais d'un cycle glottique très court, phénomène relativement courant lorsqu'une constriction glottique se relâche. Un algorithme de détection de la fréquence fondamentale par corrélation ne relèverait pas ces valeurs, puisque le coefficient de corrélation du signal à cet endroit est très bas ; en revanche, la détection individuelle des cycles par le relevé des pics de fermeture glottique, telle qu'elle est pratiquée par l'algorithme employé ici, relève ces cycles dès lors que l'amplitude des pics de fermeture en début et fin de cycle est suffisamment bien marquée ; l'algorithme est présenté dans l'Annexe 2.

Les courbes de fréquence fondamentale et de quotient ouvert des divers items représentent, à un certain niveau d'abstraction, un même phénomène linguistique, le même type de réalisation d'un accent lexical initial de mot, du fait que le contexte d'énonciation est le même. Si la même courbe se déploie sur tous les segments voisés de la syllabe, on s'attend (pour prendre un exemple) à ce que la courbe portée par /mɑːt/ *(mart)* soit proche de celle portée par le seul /ɑː/ dans /pɑːt/ *(part)*. Si en revanche la courbe ne se déploie que sur la rime, ces courbes seront dissemblables, tandis que les courbes calculées à partir de la seule rime seront similaires (*modulo* la faible perturbation locale due à l'influence de l'articulation de la consonne qui précède : sa *fréquence fondamentale co-intrinsèque*). En d'autres termes, le découpage qui révèle la plus grande proximité entre les deux ensembles de courbes est utilisé comme révélateur de la pertinence des unités linguistiques postulées par ce découpage.

La comparaison des courbes calculées sur les rimes des diverses syllabes en excluant les /m/ et /n/ initiaux montre une similarité remarquable, tandis que la courbes calculée sur l'ensemble de la partie voisée de la syllabe s'éloigne des autres. La forte différence de longueur entre les courbes avec et sans consonne initiale correspond globalement à la durée des consonnes voisées initiales /m/, /n/.

Une observation plus détaillée décèle également des différences entre les trois courbes calculées sur les seules voyelles (en vert, bleu et noir). La courbe de fréquence fondamentale après les initiales phonétiquement aspirées /p/, /t/, /k/ a une pente plus forte que celle des deux autres groupes de rimes ; la voyelle est généralement un peu plus brève. La fréquence fondamentale plus élevée en début de voisement s'explique par l'aspiration : le débit d'air étant élevé au moment où les plis vocaux s'accolent pour la mise en place de la phonation, les plis vocaux sont séparés aussitôt qu'accolés : le début de voisement se fait par un ou deux cycles très courts, peu visibles sur le signal audio (et généralement non détectés par les algorithmes de détection de la fréquence fondamentale à partir du signal audio) mais clairement identifiables sur le signal électroglottographique[1]. Les rimes de syllabes à initiale

[1] Les données présentées sur les figures 2.3 à 2.10 permettent également de mesurer la différence de fréquence fondamentale sur la voyelle selon qu'elle suit une initiale continue voisée ou une initiale non voisée, différence dont l'importance dans l'évolution des langues est bien attestée (rappelons les travaux pionniers au sujet de la tonogenèse en Asie du Sud-Est : Maspero 1912, Haudricourt 1954, 1961 ; le terme de « tonogenèse » serait dû à James Matisoff : Matisoff 1970) mais dont l'étude phonétique en synchronie demeure à approfondir (House et Fairbanks 1953, Ohala 1973, Hombert, Ohala *et al.* 1979 ; données anglaises et allemandes dans Kohler 1985 ; un passage en revue est proposé par Abramson 2004, « The plausibility of phonetic explanations of tonogenesis »). L'étude de cette question avec l'aide de l'électroglottographie paraît prometteuse dans la mesure où cette technique permet de déceler le voisement dès sa mise en place, la première fermeture glottique apparaissant distinctement sur le signal électroglottographique tandis qu'elle est peu discernable sur le signal audio. Des recherches dans ce sens sont en cours.

/m/ ou /n/ se rapprochent des rimes à initiale /b/, /d/ ou /g/ plus que des rimes à initiales /p/, /t/, /k/. Sur toutes les figures, on constate une grande proximité des courbes en noir et en vert.

Au vu des figures, et tout particulièrement de la figure 2.3, il y a lieu d'être surpris de la différence d'allure, pour les syllabes à initiale /m/ ou /n/, entre la courbe de la syllabe entière et la courbe de la seule rime. Ce phénomène tient à la procédure employée pour le calcul d'une courbe moyenne sur l'ensemble de la syllabe : la courbe représentée en rouge est calculée sans diviser la syllabe en une partie consonantique et une partie vocalique ; chaque courbe est rééchantillonnée en 100 points également espacés dans le temps, et les courbes sont ensuite moyennées pour chacun des 100 échantillons[1]. Or le rapport de durée entre consonne et voyelle n'est pas constant d'un item à l'autre ; si dans un cas la consonne /m/ ou /n/ dure 140 ms et la voyelle 100 ms, que dans un autre cas la consonne dure 80 ms et la voyelle 100 ms, le premier point de donnée de la première voyelle figurera au 58^{e} point dans le vecteur rééchantillonné, tandis que le premier point de la deuxième voyelle figurera au 44^{e} point ; ils ne seront donc pas moyennés ensemble, à la différence de ce qui se passe pour les calculs menés exclusivement sur la rime. L'allure de la courbe moyenne sur la syllabe entière n'est par conséquent pas strictement identique à celle que l'on obtiendrait en mettant bout à bout la courbe moyennée des initiales et la courbe moyennée des voyelles. Cette différence est d'autant plus visible que chaque courbe représente la moyenne d'un petit nombre d'items : 6 syllabes pour les courbes des syllabes à initiale sonante.

[1] Les raisons qui ont conduit au choix de 100 points plutôt que 5, 10 ou 20 sont exposées ci-dessous, Annexe 2, section 4.4.

Au bilan, les données apportées ici vont dans le sens de l'idée traditionnelle (exposée notamment par Pike 1948) selon laquelle la fréquence fondamentale au cours des initiales voisées ne joue pas de rôle dans la perception que les locuteurs ont des phénomènes intonatifs, et ne fait l'objet d'aucune stratégie de la part des locuteurs, la partie de la courbe coïncidant avec les segments de la rime constituant à elle seule l'unité porteuse des phénomènes prosodiques (qu'il s'agisse, selon les langues, d'un ton, d'un accent, ou de phénomènes intonatifs).

Certes, les consonnes jouent également un rôle dans l'intonation : leur allongement expressif, leur articulation plus ou moins forte constituent des indices intonatifs. De plus, sur un plan psycho-acoustique (si l'on étudie la perception des sons sur un mode non linguistique), il est possible d'entendre la partie de la courbe fondamentale qui coïncide avec une consonne continue voisée initiale telle que /l/, /m/, /n/, /ɲ/, /r/[1]. Le point de vue de Kenneth Pike, auquel les données conduisent à nous rallier, consiste à dire que la fréquence fondamentale dont les consonnes initiales sont porteuses n'a pas de rôle linguistique.

Ce constat débouche aussitôt sur de nouvelles questions. Le détail de la coordination entre gestes laryngiens et articulation segmentale appelle une étude détaillée : en fonction des consonnes et des voyelles concernées, la transition entre consonne et voyelle qui nous sert de point de repère dans la délimitation de l'initiale et de la rime consiste en des événements différents, au plan des commandes musculaires, des mouvements des articulateurs, de l'aérodynamique, et de l'acoustique. Remettant à plus tard l'étude de ces questions phonétiques, retenons l'idée d'une dichotomie entre initiale et rime : c'est la rime, et non l'ensemble de la syllabe, qui est l'unité porteuse des caractéristiques de hauteur significatives au plan linguistique.

Sur la base de ce résultat, les données naxi ont été analysées en séparant consonne initiale et rime dans les calculs, sans qu'il paraisse nécessaire de renouveler avec les données naxi la démonstration exposée ci-dessus.

1.4. Caractéristiques intrinsèques des phonèmes vocaliques

Le paragraphe qui précède illustre l'influence *co-intrinsèque* de trois catégories de phonèmes consonantiques (*lenis*, *fortis*, et continues voisées) en position initiale sur la réalisation des voyelles qui les suivent. Il est également bien établi que **les phonèmes vocaliques possèdent**

[1] Ainsi, les étudiants d'anglais qui s'exercent à écouter et reproduire les contours descendants ont parfois des doutes sur la description qui leur est fournie lorsqu'un contour descendant est porté par une syllabe à initiale continue voisée : certains, plus sensibles aux variations mélodiques à mesure de leur entraînement, entendent la partie montante portée par la consonne initiale, et en font la remarque à l'enseignant : *la courbe n'est-elle pas montante avant d'être descendante ?* (Barbara Kühnert, communication personnelle)

des *caractéristiques laryngales intrinsèques*, l'articulation supraglottique ayant une influence sur le niveau laryngal (Di Cristo 1985). La composition phonémique du corpus anglais, divisé en trois catégories vocaliques dont chacune représente un des extrêmes du trapèze vocalique (/i:/, /ɑ:/, /u:/), permet des relevés quantifiés des caractéristiques intrinsèques de ces phonèmes.

La figure 2.11 montre, pour les monosyllabes du locuteur M4 en condition S, les courbes moyennes des rimes à voyelle /ɑ:/ (ligne bleue), /i:/ (ligne verte), et /u:/ (ligne rouge), ainsi que les données brutes, sous forme de points de données (les conventions de couleurs sont identiques à celles des lignes continues). Un certain étagement se dégage pour le quotient ouvert, celui du /i:/ étant plus bas que celui du /u:/, lui-même plus bas que celui du /ɑ:/ (valeurs moyennes : 39.5 pour /i:/, 43.5 pour /u:/, 46.5 pour /ɑ:/), mais tel n'est pas le cas chez tous les locuteurs (voir en particulier les données du locuteur M7 : valeur moyenne de quotient ouvert pour /i:/ : 67.5, pour /u:/ : 63.5, pour /ɑ:/ : 64). En revanche, les valeurs de fréquence fondamentale des trois catégories vocaliques se recouvrent largement ; les points de départ et d'arrivée des courbes sont comparables, le /ɑ:/ étant légèrement plus bas dans l'ensemble. (Les valeurs moyennes sont : pour /ɑ:/ : 138 Hz ; pour /u:/ et /i:/ : 157 Hz). La voyelle /ɑ:/ se distingue des deux autres par sa longueur : 189 ms (soit 80% de plus que les deux autres : /u:/ : 104 ms, /i:/ : 101 ms), ce qui rejoint les observations de Lehiste 1970, Laver 1994 (et références citées) sur la plus grande longueur des voyelles basses. Cette différence de longueur est la seule caractéristique commune aux productions des quatre locuteurs. Les figures 2.12 à 2.14 présentent les mêmes données que précédemment pour les locuteurs M5, M6 et M7 ; chez M5 comme chez M4, /i:/ et /u:/ sont presque indistincts par leur courbe de fréquence fondamentale ; chez M7, le /u:/ a une fréquence fondamentale plus élevée ; chez M6, sa courbe est sensiblement différente de celle de /i:/, partant de plus haut et aboutissant plus bas.

La plus grande longueur de /ɑ:/ n'est pas corrélée de façon évidente à une cible finale de fréquence fondamentale plus basse : cette observation ne vaut que pour le locuteur M7. (Le nombre relativement restreint d'items ne permet pas de parvenir à des conclusions statistiquement vérifiées.) Il ne paraît pas y avoir d'inconvénient à moyenner les courbes des trois types de voyelles, considérant que les courbes plus courtes portent une variante *compressée* d'un même contour que les courbes les plus longues, plutôt qu'une version *tronquée* du contour observé sur les rimes les plus longues (au sujet de l'opposition entre troncation et compression, en anglais *truncation* et *compression*, voir Grønnum-Thorsen 1991 et Grabe, Post *et al.* 2000).

2. Résultats de mesures de débit d'air des tons vietnamiens B2, D1 et D2

L'électroglottographie, principale méthode employée dans la thèse pour l'estimation de la *qualité de voix*, ne permet pas de quantifier le débit d'air glottique ; afin de vérifier certaines des hypothèses suggérées par l'électroglottographie, une étude-pilote de débit d'air a été réalisée.

L'électroglottographie, principale méthode exploratoire employée dans la présente étude pour la caractérisation de la qualité de voix, est une mesure de la surface d'accolement des plis vocaux, non une mesure de débit d'air (onde de débit glottique). Il est facile de se représenter la relation qui existe entre longueur de la phase ouverte du cycle glottique et débit d'air glottique (*toutes choses égales par ailleurs :* à pression sous-glottique constante), relation vérifiée par l'étude de Rothenberg et Mahshie 1988, qui combine mesures de débit et électroglottographie ; quotient ouvert et débit n'en sont pas moins distincts, et ils sont obtenus par des méthodes exploratoires très différentes. Les différences de quotient ouvert observées n'ont donc pas été traduites en termes de débit d'air plus ou moins important.

Il est pourtant intéressant de poursuivre la recherche dans cette direction, pour vérifier et compléter les observations reposant sur l'électroglottographie. Notre intuition d'apprenant de langue vietnamienne (« introspection kinesthésique ») va dans le même sens que les valeurs de quotient ouvert constatées, suggérant que le débit d'air sur les rimes au ton B2 est nettement moindre que sur les rimes au ton D2. Il a donc paru utile de se donner les moyens de franchir le pas, en recourant à des mesures de débit d'air[1].

Dans une étude portant sur 6 locuteurs de dialectes du nord du Vietnam, Nguyên Van Loi et Edmondson 1998 ont caractérisé la qualité de voix des six tons apparaissant sur les rimes à sonante finale (A1 à C2) par l'emploi d'un masque de Rothenberg et un filtrage inverse des données audio à l'aide du signal de débit d'air buccal (plus précisément bucco-nasal, le masque couvrant la bouche et le nez). Ils observent, pour le ton B2 :

> All speakers start with modal voice and show progressively increasing closed-phase duration and reduction of airflow quantity over the course of the syllable, so that the general shape of the airflow in a syllable is triangular. The increasing tension finally brings the glottal folds to stasis. In this tone the airflow wave is very close to the

[1] Merci à Angélique Amelot pour ses conseils au sujet des mesures de débit.

> baseline value, which shows a high degree of passive tension. (Nguyên Van Loi et Edmondson 1998:16)

L'étude-pilote rapportée ici confirme ces observations sur le ton B2, et apporte des données quantifiées qui permettent de prolonger la comparaison avec les tons D1 et D2 proposée au chapitre III, section 3.

2.1. Déroulement de l'enregistrement

L'enregistrement, préparé avec Vũ-Ngọc Tuấn (LIMSI), a été réalisé par Bernard Roubeau à l'hôpital Tenon, service du professeur Jean Lacau Saint-Guily, le 16 mars 2004. Pour cette expérience-pilote, le consultant (M4) était l'un des enquêteurs : Vũ-Ngọc Tuấn, locuteur natif âgé de 58 ans, résident de longue date en France. L'équipement employé était la station EVA2 (voir Teston et Ghio 2002 et références citées).

Cette station permet l'enregistrement simultané du débit d'air nasal ; les séries de syllabes aux tons B2 et D2 contrastant à la fois par le ton et par la consonne finale, nasale /m/ /n/ /ŋ/ pour la série au ton B2 et occlusive /p/ /t/ /k/ pour la série au ton D2, cette mesure n'est pas dépourvue d'intérêt, et peut aider à caractériser ces syllabes. Il est néanmoins connu que le débit d'air nasal ne fournit pas une mesure directe du degré de nasalité des segments : des études aérodynamiques et fibroscopiques montrent que la durée de l'abaissement du voile du palais est toujours plus importante que le temps où le débit d'air nasal est positif (Amelot, Roubeau *et al.* 2004). Le débit d'air nasal dépend de la pression intra-orale (P_{io}) et de la résistance au niveau du port vélo-pharyngé (R_{vp}) : voir Amelot, Roubeau *et al.* 2004 et références citées. La pression intra-orale dépend de l'articulation supraglottique (réalisation des segments), en particulier du degré de constriction du conduit vocal, ainsi que de la pression sous-glottique. La relation entre degré de nasalisation acoustique et débit d'air nasal n'est pas linéaire : dès lors que le port vélo-pharyngé est ouvert, si peu que ce soit, la cavité nasale fonctionne comme un résonateur, et le son est nasalisé ; le débit d'air nasal peut être très faible, voire même négatif : l'abaissement du voile du palais peut créer une dépression d'air dans la cavité nasale (voir Amelot 2004, *Etude aérodynamique, fibroscopique, acoustique et perceptive des voyelles nasales du français*, et références citées). Une mesure plus directe du degré de nasalisation acoustique peut être obtenue à l'aide d'un transducteur ferromagnétique (ci-après désigné simplement comme *micro nasal*) placé dans l'une des narines du locuteur, à l'aide d'une olive nasale, procédé employé par Tronnier 1994, 1995,

1998 et Montagu 2004[1]. Il est envisageable de combiner cet enregistrement avec les enregistrements de débit d'air (nasal et oral), le capteur de débit d'air nasal étant placé dans une narine tandis que l'autre narine est obturée par le micro nasal. Lors de l'enregistrement-pilote à l'hôpital Tenon, le micro nasal n'a pas été employé, parce que cette expérience-pilote était consacrée principalement au débit d'air oral, et du fait des difficultés techniques que pose le branchement du micro nasal sur la station EVA2.

Au bilan, les signaux enregistrés sont : le signal audio, le signal électroglottographique (modèle de l'appareil : laryngographe de Fourcin), le débit d'air oral, et le débit d'air nasal.

Concernant les items enregistrés, comme il a été signalé en début de section 2.3, les noyaux vocaliques /wa/ et /wă/ avaient été omis de la liste de 126 syllabes qui constitue le corpus de l'expérience principale. Les syllabes correspondantes (en orthographe vietnamienne : oạm, oạn, oạnh, oạp, oạt, oạc ; oặm, oặn, oặng, oặp, oặt, oặc ; oáp, oát, oác, oắp, oắt, oắc) ont été ajoutés à la liste, portant le nombre d'items à 144 items.

Les items enregistrés sont présentés dans le tableau 32 (en orthographe vietnamienne).

1. oặc	19. ọp	37. ện	55. oạn
2. ậc	20. ợp	38. ựn	56. ác
3. ắc	21. ụng	39. ịn	57. ụm
4. oặn	22. oạc	40. ạp	58. ịp
5. ặc	23. ựng	41. iệng	59. ập
6. oáp	24. ịch	42. ạch	60. óc
7. ạn	25. ẹn	43. ạc	61. ệm
8. ợm	26. ét	44. ứp	62. úp
9. ếch	27. iệt	45. ượn	63. oạt
10. ực	28. ốt	46. oát	64. íp
11. ịnh	29. ật	47. oạnh	65. ắp
12. oắt	30. ết	48. ặt	66. oặm
13. ợt	31. ứt	49. ận	67. ướp
14. ướt	32. ít	50. ệt	68. ớp
15. oặng	33. áp	51. ạm	69. ượng
16. ộc	34. iết	52. ước	70. oác
17. ệch	35. ắt	53. oạp	71. ạnh
18. át	36. ột	54. ách	72. ếp

[1] Merci à Bernard Gautheron pour nous avoir appris cette méthode, et pour la confection d'un micro nasal pour les enregistrements de voyelles nasales du dialecte naxi de FK. Merci à Julie Montagu pour d'utiles discussions sur cette méthode.

73. ựm
74. ậm
75. ặm
76. ượm
77. ược
78. uộc
79. ộp
80. iếp
81. ép
82. iếc
83. uộng
84. ọng
85. ệp
86. ụp
87. ịm
88. ộm
89. iệm
90. ẹm
91. ốc
92. ậng
93. oặt
94. ọm
95. óp
96. ục
97. oạm
98. úc
99. ọt
100. ọn
101. oắc
102. ợn
103. ẹt
104. ượt
105. iện
106. oắp
107. ặn
108. ộn
109. ất
110. ựt
111. ịt
112. iệc
113. uốc
114. ọc
115. ạng
116. ựp
117. ốp
118. iệp
119. ẹp
120. ặng
121. ấc
122. ộng
123. ích
124. ạt
125. ệnh
126. ức
127. ót
128. ớt
129. ấp
130. ặp
131. ượp
132. *ớc
133. *ợc
134. *ợng
135. uộn
136. út
137. uộm
138. oặp
139. ụn
140. uộp
141. uột
142. uốp
143. ụt
144. uốt

Tableau 32. Corpus de syllabes vietnamiennes pour les mesures de débit d'air.

Il a été constaté que ce protocole expérimental contraignant, auquel le sujet était soumis pour la première fois, rendait difficile la concentration sur la tâche linguistique demandée : les erreurs de lecture sont nettement plus nombreuses que pour les enregistrements des locuteurs F1 et M1-M3, qui ne comprenaient qu'un enregistrement audio et électroglottographique. Un enquêteur doit suivre attentivement la lecture pour éviter les erreurs ; par ailleurs l'élocution est loin d'être naturelle. Il paraît indiqué, pour la poursuite du travail, de faire une répétition de la séance à l'avance.

Le masque qui couvre la bouche a pour effet de rendre la respiration plus difficile, ce qui peut susciter une réaction d'*hyper-ventilation* : le sujet a tendance à respirer à un rythme accéléré, en réaction à la sensation d'étouffement produite par le masque[1]. Ce comportement peut perturber l'élocution. Le sujet (qui dans cette expérience-pilote était également l'un des enquêteurs) était prévenu du fait, et s'est efforcé d'éviter le réflexe d'hyper-ventilation.

Les fichiers créés en sortie par la station EVA2, codés dans un format propriétaire, ont été transformés en fichiers .wav (lisibles par les logiciels de traitement et d'analyse du signal) à l'aide de scripts MATLAB aimablement communiqués par Shinji Maeda (ENST).

La durée de chaque série d'items est limitée par le fait que l'étalonnage de la station EVA2 perd en précision en cours d'enregistrement.

Six séries ont été enregistrées :

série 1 : rimes 1 à 84 sans initiale, hors cadre. (Il y a eu une erreur sur la syllabe 76.)

série 2 : rimes 85 à 144 sans initiale, hors cadre. En début d'enregistrement, le locuteur a lu plusieurs tons B2 comme des tons A2, ce qui a nécessité un nouveau départ. Il paraît y avoir une inexactitude dans la réalisation des phonèmes de 115 et 116, mais la déformation du signal audio par le masque servant à mesurer le débit d'air oral ne permet pas d'être sûr du fait.

séries 3 et 4 (rassemblées dans un même fichier) : rimes 1 à 144 avec initiale /t/, hors cadre.

série 5 : rimes 1 à 58 sans initiale, en phrase-cadre I.

série 6 : rimes 1 à 58 sans initiale, en phrase-cadre R.

Le nombre total de syllabes correctement lues (c'est-à-dire sans confusion sur la voyelle ni le ton) est de 388.

---

[1] Merci à Bernard Gautheron pour nous avoir signalé à l'avance cet écueil, et plus généralement pour ses conseils de praticien au sujet des enregistrements électroglottographiques et de débit d'air.

2.2. Pose des bornes de début et fin

Il serait commode de déterminer les bornes de début et de fin de chacune des rimes à partir du signal électroglottographique, comme cela a été fait pour les mesures d'intensité acoustique présentées ci-dessus (pour l'anglais, le naxi et le vietnamien) : se fonder sur les première et la dernière fermetures glottiques. Cela n'a pas paru pratiquable pour ces données, car le signal électroglottographique présente des irrégularités relativement nombreuses, qui s'étendent sur plusieurs périodes ; le signal électroglottographique n'a finalement pas été utilisé. Les bornes de début et de fin ont finalement été posées au vu du signal audio, selon un critère d'intensité : les points relevés sont ceux où le signal dépasse la barre des –32 dB, seuil choisi empiriquement suite à l'examen du signal, et utilisé pour toutes les séries (1 à 6). Ce seuil relativement élevé revient à sélectionner une portion de signal relativement brève (que l'on peut désigner comme le cœur de la rime). Un script MATLAB relève la position et l'amplitude des maxima locaux de débit d'air buccal et de débit d'air nasal, et fournit les valeurs moyennes et médianes de ces paramètres ; une courbe moyenne est également calculée sur toute la longueur des intervalles retenus.

2.3. Résultats de débit d'air oral

A la différence du pneumotachographe de Rothenberg (Rothenberg 1977), dont la résolution temporelle est de ½ ms, ce qui permet de distinguer le passage d'air lors de chacun des cycles glottiques, la station EVA2 donne une courbe relativement lisse sur la syllabe ; seules de légères fluctuations locales permettent parfois de détecter la périodicité. Les courbes correspondent à l'allure générale de l'évolution du débit d'air sur la syllabe, de sorte qu'il apparaît possible de les moyenner entre elles sans perte d'information

Les figures 3.1 à 3.4 présentent, pour les quatre conditions de lecture (R ; I ; isolé sans consonne initiale ; isolé avec consonne initiale /t/) les courbes de débit d'air oral des tons B2, D1 et D2 en 20 points également espacés dans le temps, la longueur de chacune des courbes étant égale à la durée moyenne des intervalles retenus pour l'analyse (approximation de la durée syllabique).

Ces figures montrent que le débit d'air des tons D1 et D2 est comparable à la fois quantitativement et qualitativement (dans son allure générale au cours de la rime), et se différencie de celui du ton B2, globalement plus bas et ne présentant pas régulièrement de montée finale. Le débit d'air des tons D1 et D2 s'élève avec régularité au cours de la rime.

Les figures 3.5 à 3.7 comparent, ton par ton, les réalisations d'une condition de lecture à l'autre. Le ton D1 est peu modifié ; le ton D2 a un débit nettement plus haut en fin de syllabe ; le ton B2 a un débit légèrement modifié, l'allure de la courbe devient montante, et

son débit global paraît diminuer. Le ton D2 a un débit qui croît très rapidement en cours de syllabe en condition I : il dépasse le ton D1 en fin de syllabe.

Les résultats sont présentés dans le tableau 33.

condition	ton	position moyenne du pic de dab (sur 100 points)[1] ; entre parenthèses : valeur médiane	maximum moyen du pic de dab, en ml/s ; entre parenthèses : valeur médiane
R (51 syllabes)	B2	50 (53.5)	136 (142)
	D1	87 (89)	332 (305)
	D2	82 (84.5)	288 (250)
I (54 syllabes)	B2	76 (79.5)	117 (117)
	D1	88 (91)	368 (366)
	D2	83 (85)	422 (404)

Tableau 33. Valeurs de débit d'air oral pour les syllabes en phrase-cadre

La variabilité de la position du pic de débit d'air oral au ton B2 (vers 80 en I, vers 50 en R) peut être attribuée au caractère relativement plat de la courbe.

L'opposition entre les tons B2 et D2 est réalisée de façon sensiblement plus nette en condition I qu'en condition R : en condition R, le maximum de débit atteint au ton D2 est environ deux fois plus élevé en moyenne qu'au ton B2 ; en condition I le maximum est 3.5 fois plus élevé pour le ton D2. Ces observations confirment ce qu'indiquaient les données électroglottographiques : la qualité de voix des tons B2 et D2 a tendance à être plus différente encore en condition I qu'en condition S.

Les valeurs obtenues sur les syllabes hors cadre sont à interpréter avec prudence : pauses, respirations, début de la phonation influencent le débit observé sur les syllabes-cible. Les valeurs obtenues sont les suivantes :

[1] Les raisons qui ont conduit au choix de 100 points plutôt que 5, 10 ou 20 sont exposées ci-dessous, Annexe 2, section 4.4.

condition	ton	position moyenne du pic de dab (sur 100 points) ; entre parenthèses : valeur médiane	maximum moyen du pic de dab, en ml/s ; entre parenthèses : valeur médiane
hors cadre, sans initiale (141 syllabes)	B2	57.5 (52.5)	203 (206)
	D1	83.2 (86.5)	403 (412)
	D2	82 (87)	427 (439)
hors cadre, avec initiale /t/ (142 syllabes)	B2	17 (2)	243 (243)
	D1	42 (33)	391 (396)
	D2	50 (79)	375 (367)

ton	moyenne de dab en condition R, en cm^3/s	moyenne de dab en condition I, en cm^3/s	significativité des différences de dab entre conditions
B2	58.0	93.3	1-62
D2	174	213	aucun

Tableau 34. Valeurs de débit d'air oral (abrégé en dab*) pour les syllabes isolées*

Ces données confirment :

- que le pic est plus tardif au ton D2 qu'au ton B2
- que le pic de débit d'air est plus élevé au ton D2 qu'au ton B2
- que les valeurs des tons D1 et D2 sont d'un ordre de grandeur comparable.

Ces données montrent aussi qu'après la consonne initiale non voisée /t/, le pic de débit est déplacé : il intervient plus tôt. Pour le ton D2 le maximum demeure le plus souvent en fin de syllabe.

Les valeurs de débit d'air des tons D1 et D2, surtout en fin de syllabe, se rapprochent des valeurs indiquées par Henton, Ladefoged *et al.* 1992 pour la voix soufflée *(breathy voice)* : ces auteurs rapportent les valeurs de débit d'air suivantes pour un sujet mâle, et un niveau de pression sous-glottique de 8 cm d'eau :

type de phonation	débit d'air (en ml/s)
aspiration des consonnes aspirées	1000
voix soufflée *(breathy voice)*	500
voix modale	120
voix tendue *(stiff voice)*	100
voix craquée/laryngalisation	< 100

Tableau 35. Valeurs typiques de débit d'air oral pour 5 modes phonatoires, d'après Henton, Ladefoged et al. 1992.

Catford 1977:99 indique pour la voix soufflée *(breathy voice)* des valeurs de débit d'air plus basses, de l'ordre de 300 à 400 cm^3/s ; rapportées à ces valeurs, les observations du ton D2 du locuteur M4 sont dans la plage de la voix soufflée.

Enfin, l'écart-type observé dans les mesures de débit d'air est beaucoup plus élevé que celui observé dans les mesures de quotient ouvert des locuteurs M1-M2-M3-F1. Si la variabilité inter-syllabes des mesures de débit se confirmait pour un plus grand nombre de locuteur, elle indiquerait

- que l'articulation supraglottique (identité phonémique de la voyelle et de la consonne) a un effet plus important sur le débit d'air que sur la qualité de voix
- que les tons vietnamiens B2, D1 et D2 sont définis plus précisément en termes de configuration laryngale qu'en termes de débit d'air.

Les résultats des mesures de débit d'air nasal, qui ne sont pas directement pertinentes pour la présente étude, sont détaillés dans un article soumis pour publication, en collaboration avec Vu-Ngoc Tuân, Angélique Amelot et Bernard Roubeau.

3. Etude de DECPA, amplitude des pics de fermeture sur le signal électroglottographique

La recherche de paramètres complémentaires de F_0 et du quotient ouvert a été motivée par la recherche de corrélats de l'*effort vocal*. Nos premières mesures de quotient ouvert s'inscrivaient dans le cadre d'une étude du marquage intonatif du focus[1]. Dans de nombreuses langues dont l'anglais, le vietnamien et le naxi, l'importance relative des syllabes au plan de la hiérarchie de l'information se reflète dans l'intonation ; il paraissait intéressant de réaliser une approche expérimentale des mécanismes par lesquels se réalisent divers degrés de mise en valeur. Or **les premières mesures n'ont pas fait ressortir de**

[1] Cette étude est rapportée dans notre contribution à l'ouvrage *Fonction et moyens d'expression de la focalisation à travers les langues,* Mémoires de la Société de Linguistique de Paris, 2003 (Michaud 2003) ; les principaux résultats sont repris et mis en perspective ci-dessous, chapitre II, 2e partie, section 2.

corrélation négative nette entre quotient ouvert et mise en valeur intonative d'un mot. Une baisse de quotient ouvert apparaissait comme un des corrélats d'une *réalisation soignée* des syllabes ; en revanche, le quotient ouvert était relativement élevé dans les cas d'*accent d'insistance*, dans lesquels il nous semblait précisément percevoir l'effort vocal le plus fort.

Sur le conseil de Nathalie Henrich, l'investigation a été étendue à un autre paramètre calculé à partir du signal électroglottographique : l'amplitude des pics de fermeture, ou, en anglais, DECPA, *Derivative-Electroglottographic Closure Peak Amplitude*[1]. Ce paramètre est présenté ci-dessous, reprenant l'essentiel d'un exposé à la conférence *Speech Prosody 2004* : « A Measurement from Electroglottography: DECPA, and its Application in Prosody » (Michaud 2004).

La section 1 représente une introduction à la question générale de la mesure de la pente spectrale ; la section 2 est spécifiquement consacrée à DECPA.

3.1. Introduction : la question de la mesure d'intensité, et l'utilité de mesures de la pente spectrale

A l'intérieur de la triade F_0-durée-intensité, l'intensité est le parent pauvre. Sa mesure n'est pas interprétable de façon transparente, pour les raisons suivantes :

- l'intensité des différents segments (et notamment des différentes voyelles) n'est pas directement comparable : chaque phonème possède une intensité intrinsèque ; l'intensité d'un segment est en outre influencée par son contexte phonémique, par un phénomène d'intensité *co-intrinsèque* (Rossi 1971, Fletcher 1972, et références citées, dont des travaux de Kenneth Stevens)
- l'intensité varie avec la distance entre le micro et la bouche du locuteur. Les variations d'intensité peuvent être dues pour une part non négligeable aux changements de position et d'orientation de la tête du locuteur (l'emploi d'un micro-casque permet de pallier cet inconvénient, mais peut en présenter d'autres ; voir ci-dessous, section 2.2.1).

---

[1] Le nom DECPA a été donné par Nathalie Henrich, qui nous a signalé l'intérêt potentiel de ce paramètre, déjà mentionné, dans l'étude de la voix chantée, par Bertrand Chuberre (*Les registres et passages dans la voix chantée*, Chuberre 2000:131-135). La première mention d'une mesure de l'amplitude des pics sur le signal DEGG dans les travaux publiés semble être chez Dromey, Stathopoulos *et al.* 1992 ; cependant, ces auteurs considèrent, non la valeur des pics en elle-même, mais le rapport *amplitude du pic de fermeture/amplitude du pic d'ouverture*. Ils nomment cette valeur SQ_{dEGG}, c'est-à-dire *quotient de rapidité calculé à partir du signal dérivé du signal électroglottographique* (p. 48). Le terme de « quotient de rapidité » ne paraît en fait pas adapté pour cette mesure (Nathalie Henrich, communication personnelle).

En outre, des études perceptives déjà anciennes paraissent démontrer que l'oreille est peu sensible aux changements d'intensité, en comparaison de sa sensibilité aux variations de hauteur. Enfin, en anglais, langue la plus étudiée, allongement, augmentation d'intensité et mouvements de F_0 se concentrent dans une large mesure sur la syllabe porteuse de l'accent lexical, ce qui a pu encourager à considérer que ces trois dimensions étaient largement redondantes[1] (en français, la corrélation est moins nette : par exemple, la *continuation* se traduit par une montée de F_0 souvent accompagnée d'une décroissance d'intensité [Delattre 1966]).

Mais **les expériences qui avaient conduit, à partir des années 1950, à prêter à l'intensité du son un rôle linguistique très secondaire contenaient en fait une erreur de méthode.** Dans les expériences de manipulation de l'intensité menées par Fry 1955, 1958, Mol et Uhlenbeck 1956, Issatchenko et Schädlich 1966, l'intensité était modifiée par un changement de gain global, qui affectait également toutes les fréquences ; or dans la parole naturelle, l'augmentation de l'intensité vocale résulte de phénomènes plus complexes. Il est nécessaire de tenir compte de cette spécificité, celle de la présence ou absence d'*effort vocal*. Au plan perceptif, l'oreille sait reconnaître cet effort : un stimulus produit avec un effort vocal plus grand sera jugé plus intense, même si l'intensité globale est en fait identique dans les deux cas (voir Glave et Rietveld 1975 et références dans Vaissière 2004). « L'oreille opère une sommation des composantes spectrales différente de celle qui régit la mesure de l'intensité globale » (Rossi 1971:143).

Etudiant le néerlandais et l'anglais américain, Sluijter, Van Heuven et Pacilly montrent le rôle perceptif de la répartition spectrale de l'intensité (ou *pente spectrale*), découverte qui relance la question des paramètres prosodiques. Au sujet du rôle linguistique de la pente spectrale, ces auteurs apportent une réponse tranchée : beaucoup d'intensité dans les hautes fréquences (« high-frequency emphasis ») serait un corrélat de l'accent lexical (en anglais : *stress*), dont un autre corrélat important est la durée, tandis que les mouvements de F_0 seraient des corrélats de l'insistance intonative et de la focalisation contrastive (Sluijter, Van Heuven *et al.* 1997, « Spectral balance as a cue in the perception of linguistic stress »).

Au plan linguistique, cette dichotomie est surprenante : l'expérience des recherches en prosodie semble au contraire montrer qu'il n'existe pas de correspondance terme à terme entre faits linguistiques et paramètres acoustiques, ces derniers étant liés, tant au plan de la production qu'au plan de la perception, par des relations de compensation (en particulier entre hauteur, intensité, débit, et rapports de durée, entre consonnes et voyelles et entre syllabes). Il ne manque pas d'observations qui contredisent la proposition de Sluijter, Van Heuven et Pacilly : F_0 est un des corrélats connus de l'accent lexical ; parallèlement, F_0 n'est

[1] Voir les travaux de Wayne Lea en reconnaissance automatique de la parole (depuis Lea 1973).

pas le seul corrélat des morphèmes d'intonation pragmatique : pour le suédois, Christer Gobl montre que l'intensification intonative est associée à un surcroît d'énergie dans les hautes fréquences (Gobl 1988) ; Gunnar Fant et Anita Kruckenberg décrivent, également en suédois, des différences spectrales sous l'accent lexical, mais précisent que ces différences sont plus marquées sur les mots qui reçoivent un accent d'insistance (Fant et Kruckenberg 1995:623). De même, M. Heldner utilise la mesure de pente spectrale pour détecter les phénomènes d'ordre pragmatique qui selon Sluijter, Van Heuven et Pacilly se réaliseraient uniquement par des variations de F_0 (Heldner, Strangert *et al.* 1999 ; Heldner 2001a, 2001b, 2003). En anglais, N. Campbell observe une augmentation de l'énergie dans les hautes fréquences lorsqu'il y a mise en valeur (Campbell 1995). **Ces divers résultats confirment l'importance de la pente spectrale comme paramètre intonatif, mais établissent qu'elle n'a pas pour seule fonction de signaler l'accent lexical.**

Au plan expérimental, la mesure de la pente spectrale est problématique. L'amplification des hautes fréquences est liée à la forme de l'onde de débit glottique (une introduction à cette question se trouve dans Titze 1994:117-120). **Il n'existe actuellement pas de méthode établie pour la mesure de l'onde de débit glottique et de la pente spectrale.** Il n'existe pas encore de méthode éprouvée pour relier les signaux électroglottographiques à la pente spectrale ; l'idée d'observer l'amplitude des pics de fermeture sur la dérivée du signal électroglottographique repose en revanche sur un fait certain : une plus grande rapidité de fermeture de la glotte a pour effet une pente spectrale moins forte, c'est-à-dire une plus forte amplitude des harmoniques élevés (Colton 1984). Cette rapidité peut s'exprimer par le quotient de vitesse, rapport entre la durée de la phase d'ouverture et la durée de la phase de fermeture. Diverses méthodes d'approximation de ce paramètre à partir du signal électroglottographique ont été proposées, mais la détection des points nécessaires à ce calcul, point d'ouverture maximale et point de fermeture maximale, repose sur le choix empirique d'un seuil ; or la forme des signaux varie beaucoup, de sorte que la comparaison (même à l'intérieur d'un même enregistrement) est hypothéquée d'emblée. En revanche, il est possible de mesurer de façon précise l'amplitude du pic de fermeture sur le signal dérivé, qui donne la pente maximum atteinte par le signal électroglottographique lors de la fermeture de la glotte (voir, à nouveau, la figure 1.6). Cette indication sur la rapidité de fermeture des plis vocaux paraissait avoir toutes chances d'être liée à la pente spectrale du signal audio.

3.2. A quoi correspond le paramètre DECPA ?

Il est établi que la pente spectrale est inversement proportionnelle à la rapidité de la fermeture glottique. **La mesure de DECPA correspond à un seul point par période : elle correspond à la vitesse maximale d'augmentation du contact des plis vocaux atteinte au cours de la fermeture glottique.** DECPA est mesurable même sur des enregistrements

électroglottographiques relativement bruités (en particulier dans les enregistrements de voix féminines) : l'amplitude des pics de fermeture est toujours supérieure à celle des pics d'ouverture en mécanisme 1 (« voix de poitrine » ; au sujet des mécanismes phonatoires, voir l'entrée correspondante du Glossaire), et est également supérieure en mécanisme 2 dans la plupart des cas.

DECPA varie fortement au cours d'un même énoncé (dans une proportion de 1 à 5 ou plus) ; sa valeur absolue n'a pas de signification en elle-même, puisqu'elle dépend du niveau d'amplification du signal choisi lors de l'enregistrement. Son étude doit donc se faire en valeurs relatives, comparant les valeurs pour un même locuteur au cours d'une même séance d'enregistrement ; la mesure de DECPA peut par exemple servir à comparer diverses syllabes au sein d'un même énoncé, ou divers ensembles de syllabes au cours d'un même enregistrement.

Cette mesure est employée à titre accessoire dans l'étude-pilote consacrée à l'intonation pragmatique en naxi (chapitre II, 2^{e} volet), et dans l'étude des tons vietnamiens B2 et D2 (chapitre III, section 3.3.4). Les courbes obtenues ne peuvent être traduites directement en termes d'intensité acoustique, de débit glottique ou de pente spectrale. Elles peuvent néanmoins appeler l'attention vers certaines caractéristiques du signal électroglottographique : ainsi, dans le cas des tons vietnamiens B2 et D2, les figures 3.79 et 3.80 (chapitre III) montrent une divergence entre l'allure des courbes aux tons B2 et D2 qui se situe essentiellement sur la deuxième moitié de la syllabe. Les mesures de débit d'air rapportées dans l'Annexe 2 (section 2) révèlent un débit d'air élevé en fin de syllabe au ton D2, ensemble de données qui est caractérisé par une courbe ascendante de DECPA. Dans le cas de la voix masculine en mécanisme 1, il semblerait donc que la mesure de DECPA permette de formuler des hypothèses concernant le débit d'air : DECPA paraît positivement corrélé au débit d'air. Le degré de corrélation entre ces deux paramètres, qui varie certainement avec le mécanisme phonatoire, n'a pas été étudié ici de façon statistique faute d'une quantité suffisante de données de débit d'air et de la qualité non optimale de l'enregistrement électroglottographique réalisé simultanément avec les mesures de débit d'air.

Dans l'ensemble, il apparaît que DECPA ne représente pas l'approximation de la pente spectrale à la source qui en était escomptée. L'interprétation de DECPA est malaisée, du fait que

- le changement global d'amplitude du signal électroglottographique peut être dû à plusieurs facteurs, dont la position des électrodes ; un mouvement vertical du larynx a très vraisemblablement une incidence sur DECPA
- les pics de fermeture doubles sont relativement courants ; quel pic choisir dans la mesure de DECPA pour le cycle glottique considéré : le premier pic, le plus ample des deux, ou

le dernier ? Il semble plutôt que dans ces cas, la mesure de DECPA ne soit pas applicable, ce qui réduit beaucoup son champ d'application.

La comparaison des courbes de DECPA et de débit d'air correspondant aux tons vietnamiens B2, D1 et D2 montre une certaine similarité entre les courbes de DECPA du locuteur M1 présentées sur les figures 3.79 et 3.80 (chapitre III) et les courbes de débit d'air du locuteur M4 (Annexe 2, section 2). Cela suggère une utilisation possible du paramètre DECPA : il permettrait de formuler des hypothèses concernant le débit d'air, DECPA étant corrélé au débit d'air. Le degré de corrélation entre ces deux paramètres, qui varie certainement avec le mécanisme phonatoire, n'a pas été étudié ici de façon statistique faute d'une quantité suffisante de données de débit d'air. (Une étude portant sur un nombre accru de locuteurs est en cours.)

4. Scripts développés dans l'environnement MATLAB[1]

Les principaux scripts développés pour le propos de la présente thèse sont les suivants :

Nom	Type de script	Cahier des charges
aver	interpolation et moyennes	création d'une courbe moyenne et d'une simulation de syllabe à partir d'un ensemble de mesures séparées. Voir section 1.4.
beginend	importation de données	lire un fichier texte contenant une annotation créée à l'aide du logiciel SoundForge (« Regions List »), et fournir en sortie les noms des balises et les temps de début et fin (en millisecondes).
correctread	importation de données	corriger les erreurs de la fonction d'importation de MATLAB dlmread.m
forver	interface de mesure semi-automatique de formants	appeler des algorithmes de détection de formants, afficher les résultats sur spectrogrammes, et permettre une vérification par l'utilisateur. Voir section 2.
loadpraat	importation de données	importation de fichiers de résultat au format .txt créés par Praat *(short text file)*
pairedttest2	statistiques	réaliser un test-t apparié et fournir les valeurs de t et p. Voir section 1.3.
peakdet	interface de mesures semi-automatiques sur le signal électroglottographique	Calculer F_0, quotient ouvert et DECPA à partir d'un signal électroglottographique. Voir section 1.1.

[1] Cette Annexe ne détaille pas les scripts réalisés par d'autres chercheurs, et utilisés dans la présente étude : ainsi, les données de la station EVA2 sont convertis par des scripts fournis par Shinji Maeda ; les spectrogrammes affichés par <forver> sont produits par le script <myspecgram>, écrit par Paul Kienzle et modifié par Sean Fulop ; la gestion des fenêtres par <peakdet> recourt au script <tilefigs>, créé par Charles Plum.

4.1. Outil d'analyse du signal électroglottographique : implémentation logicielle de la méthode de détection des pics positifs et négatifs sur la dérivée du signal

Suite au choix de suivre la méthode de calcul du quotient ouvert (et de la fréquence fondamentale) par l'utilisation de la dérivée du signal électroglottographique, se posait la question des outils logiciels pour appliquer cette méthode.

4.1.1. Programmes existants au début de notre recherche

Au moment où la présente étude a commencé, aucun outil permettant ce type d'analyse n'était distribué (ni logiciel librement disponible, ni logiciel commercial)[1]. C'est grâce à l'aide de chercheurs qui avaient programmé des algorithmes remplissant cette fonction que l'étude a commencé.

4.1.1.1. Programmes de Nathalie Henrich

Nathalie Henrich (LAM) nous a aimablement communiqué ses programmes écrits dans l'environnement MATLAB, programmes dont l'algorithme est présenté en détail dans Henrich 2001 et Henrich, d'Alessandro *et al.* 2004 : il est appelé DECOM, pour *Derivative-Electroglottographic Correlation-based method for Open quotient Measurement.* **Cet algorithme est conçu pour des extraits de voix chantée, dans un registre relativement précis** : voix de chanteurs professionnels ou semi-professionnels, tous de formation classique, pour la plupart issus du Conservatoire National Supérieur de Musique et de Danse de Paris. Fonctionnant par corrélation sur le signal dérivé, il fournit une valeur d'une grande précision dans les cas où le signal donné en entrée respecte le critère de quasi-périodicité (précision d'1 Hz pour la mesure de F0) ; en revanche, **si le critère de quasi-périodicité n'est pas respecté, le programme considère que le passage en question n'est pas voisé. Or la voix**

---

[1] Ce vide est en partie comblé, depuis 2004, par le logiciel Speech Filing System (http://www.phon.ucl.ac.uk/resource/sfs/) distribué librement par University College London. Ce logiciel n'est pas spécifiquement dédié à l'analyse électroglottographique mais propose une fonction d'estimation des durées respectives de la phase ouverte et de la phase fermée du cycle glottal (fonction nommée TGXGEN). L'algorithme est le suivant :

1) Low-pass filter the Lx waveform at 3000Hz.

2) Differentiate the Lx waveform

3) Find maxima and minima of differentiated signal that lie above a threshold. These are taken as the positions of closure and opening on the lx waveform

4) The positions of larynx opening are determined as the next point in the lx waveform at amplitude equal to that at closure.

N'ayant eu connaissance de cet outil qu'en décembre 2004, alors que nos propres outils étaient déjà au point, nous n'y avons pas eu recours.

parlée connaît des changement rapides de fréquence fondamentale ; en termes de qualité de la voix, le changement peut également être rapide et de grande ampleur, de sorte que la durée des cycles ainsi que la forme du signal électroglottographique varient d'un cycle à l'autre dans des proportions bien supérieures à ce qui s'observe dans le chant classique.

La répétition à l'identique *stricto sensu* n'existe pas dans le domaine de la voix humaine, qui ne saurait donc être exactement périodique. Le niveau d'exigence par lequel est définie la quasi-périodicité peut être varié en fonction des signaux auxquels il est appliqué ; mais pour sortir du domaine de la périodicité au sens strict on n'entre pas pour autant dans le domaine de l'arbitraire : il existe des critères mathématiques pour évaluer le degré de périodicité d'un signal. Il n'aurait pas été avisé de jouer sur les coefficients pour appliquer à la parole l'algorithme DECOM développé pour le chant : **de nombreux passages que nous souhaitions analyser ne satisfont pas au critère de quasi-périodicité tel qu'il est employé en acoustique.**

Une autre limite de l'algorithme DECOM est qu'il n'est pas conçu pour fournir des valeurs pour les premières et dernières périodes de voisement *(voice onset, voice offset)*, alors que le passage très fréquent du voisement à l'absence de voisement au cours de la voix parlée rend cruciale la prise en compte de ces cycles dans l'analyse linguistique.

4.1.1.2. Programmes de Vu-Ngoc Tuân

Parallèlement, Vu-Ngoc Tuân (LIMSI) avait développé un outil ne recourant pas à la corrélation : algorithme fonctionnant par **détection des pics positifs et négatifs sur la dérivée du signal électroglottographique, sans poser de condition sur la périodicité du signal**. Ces programmes sont utilisés par Vu-Ngoc, d'Alessandro *et al.* 2002 dans la description des tons du vietnamien. Vu-Ngoc Tuân a compilé ses programmes de façon à ce qu'ils puissent fonctionner dans l'environnement Windows, nous a guidé dans leur utilisation, et a réalisé, à mesure des problèmes rencontrés, des modifications progressives des programmes, modifications dont il est intéressant de détailler certaines étapes. La première version de l'algorithme était conçue de la façon suivante :

- chargement du signal électroglottographique
- dérivation
- séparation en deux partie : partie positive (utilisée pour la détection des pics de fermeture) et partie négative (pour la détection des pics d'ouverture)
- découpage des signaux résultants en cadres de 2048 points ; calcul de l'énergie de chaque cadre ; calcul du maximum pour la rime traitée, nommé $<w_{max}>$, et du minimum, $<w_{min}>$ (en dB) ; définition d'un seuil d'énergie $<seuil(w)>$, par la formule suivante : $seuil(w) = w_{min} + 0.05 * (w_{max} - w_{min})$

- somme des amplitudes à l'intérieur des cadres dont l'énergie est supérieure à <seuil(w)> ; en est tirée l'amplitude moyenne, <amin>, moyenne de tous les cadres non exclus ; <amin> sert ensuite de seuil d'amplitude.
- seuillage du signal, ne retenant que les valeurs supérieures à <amin>. Le signal résultant est appelé $DEGG_S$.

- pour chacun des points de $DEGG_S$, calcul de la somme des amplitudes des 11 points qui l'entourent (5 à gauche, 5 à droite, plus le point lui-même). Cela donne DDW, le signal lissé. DDW est donc constitué d'une série de bosses positives (ou négatives pour les ouvertures).

- Calcul de la somme des amplitudes <WP> dans chacune des bosses.
- Exclusion des <WP> qui se trouvent à moins de 10% de la valeur la plus élevée de <WP>.
- Détermination des bords gauche et droit des <WP> restants : LE(i), RE(i).
- Détermination des pics en retournant au signal dérivé non lissé pour rechercher le maximum situé entre LE(i) et RE(i). (Par la suite, les maxima ont été recherchés sur le signal lissé, non sur le signal non lissé original, ce dernier présentant un certain bruit de fond.)

Cette première version des programmes faisait subir deux seuillages au signal. La procédure a ensuite été simplifiée comme suit :

- dériver le signal
- déterminer un seuil pour la partie positive et un autre pour la partie négative
- lisser le signal
- déterminer les régions de pics en appliquant le seuil
- ne conserver que les régions ayant une aire suffisamment grande.

La version la plus récente des programmes, en langage MATLAB, accompagnée d'une documentation, est disponible à l'adresse http://www.lam.jussieu.fr/src/Membres/Henrich/egg/. Les paragraphes qui suivent présentent les orientations générales du développement logiciel réalisé (largement dues à Vu-Ngoc Tuân).

4.1.2. Orientations du développement logiciel effectué au cours de la présente étude

4.1.2.1. Choix d'une méthode semi-automatique

Une constante de la méthode employée dans la présente recherche est son caractère semi-automatique : pour chaque rime traitée, les résultats obtenus par application de l'algorithme sont affichés pour que l'utilisateur puisse les confronter avec le passage correspondant sur le signal électroglottographique et son signal dérivé, et le cas échéant supprimer les valeurs erronées ou modifier les coefficients employés dans le calcul afin de détecter tous les cycles glottiques.

Une méthode entièrement automatique pourrait paraître gage d'une plus grande objectivité ; elle serait également moins coûteuse en temps. La méthode semi-automatique comporte pourtant plusieurs avantages. Le paramétrage de l'outil se fait de façon empirique, dans la méthode automatique comme dans la méthode semi-automatique ; il est préférable de conserver le loisir de l'ajuster à chaque item traité plutôt que de le fixer par avance pour tous les items. Quant au temps que demande le traitement semi-automatique des données (une petite demi-journée de travail pour calculer les valeurs de fréquence fondamentale et quotient ouvert d'une centaine de syllabes), il paraît moins considérable lorsqu'il est mis en regard des moyens déployés pour l'expérience dans son ensemble : en amont, la préparation de l'enregistrement, sa réalisation technique, le transfert du signal vers un support informatique, la pose manuelle des balises de début et de fin des intervalles à analyser ; en aval, le traitement statistique des données et la réflexion auquel elles donnent lieu. **Au vu de l'ampleur de l'entreprise, le temps consacré au bon établissement des données, par leur vérification manuelle, paraît un investissement de temps raisonnable**, d'autant plus qu'il fournit l'occasion d'observer le détail des données.

La méthode semi-automatique est permise par certains logiciels, notamment XWaves (qui n'est plus maintenu ni distribué, ce qui contraint progressivement ses utilisateurs de passer à d'autres outils), pour la mesure de formants ; elle est notamment employée dans une étude récente de Adank, van Hout *et al.* 2004 (avec un outil créé pour le propos de leur recherche).

4.1.2.2. Choix de ne pas appliquer de traitement sur le signal enregistré

Le signal enregistré n'est soumis à aucun traitement préalable à son analyse : ni filtrage des basses fréquences pour supprimer les mouvements verticaux (ce que Fourcin appelle la composante *Gx* ; voir Fourcin 1971), ni filtrage des hautes fréquences pour supprimer les irrégularités du signal lorsqu'il est de qualité médiocre. En effet, tout filtrage revient à supprimer une information, supposée inutile dans la perspective d'une certaine utilisation,

mais qui peut éventuellement être informative pour un autre type d'analyse. Si lissage il y a, il est effectué *en ligne*, pendant l'analyse.

4.1.3. Principes pour la détection des pics de fermeture

Les pics positifs sont repérés par une méthode de seuil relatif appliquée à la dérivée du signal électroglottographique. **Le choix du seuil a varié d'une version du programme à l'autre, sa définition étant adaptée en fonction des signaux à traiter** : l'essentiel est que le seuil choisi permette de déceler chacun des pics qui marquent le début d'un cycle glottique, tout en excluant les plus petites variations du signal électroglottographique qui pourraient se confondre avec des pics de fermeture.

Un seuil à 25% du maximum local (valeur maximale du signal dérivé sur l'intervalle étudié) a pu paraître adapté ; mais la variation d'amplitude peut être considérable à l'intérieur d'une même rime (une augmentation très rapide d'amplitude est par exemple observée au cours des rimes au ton D2 du vietnamien), de sorte que le seuil à 25% a pour effet d'écarter certaines périodes.

Au vu de ces difficultés, Vu-Ngoc Tuân a proposé de choisir pour seuil l'opposé de la valeur minimale sur l'intervalle, ce qui revient à choisir pour seuil l'amplitude du pic d'ouverture le plus saillant. Cette définition donne des résultats satisfaisants dans le cas du mécanisme phonatoire 1, dans lequel l'amplitude des pics d'ouverture est nettement inférieure à celle des pics de fermeture. En revanche, en mécanisme 2, l'amplitude des deux types de pics devient similaire ; dans certains cas, le seuil ainsi fixé aboutit à ce que plusieurs des pics de fermeture ne soient plus détectés[1].

Le choix finalement effectué dans le programme <peakdet> consiste à placer le seuil à 50% du maximum du signal dérivé, et à permettre à l'utilisateur de le modifier :

- en lui assignant une valeur absolue (au vu du signal dérivé)

---

[1] Enfin, sur certains signaux rencontrés (langue naxi, locuteur M5), le fait de fixer un seuil unique pour l'ensemble de la rime aboutit, soit à exclure les fermetures de certains cycles en fin de rime, soit à considérer comme pics de fermeture des remontées entre deux fermetures d'assez grande ampleur (sur la première moitié de la rime) qui ne sauraient correspondre physiologiquement à des fermetures, et que nous expliquons comme des artefacts d'enregistrement dûs au filtrage (filtre passe-haut) effectué par le glottographe EG2 : une composante de basse fréquence corrigée par les circuits électroniques de l'appareil (dont la documentation est peu fournie) peut avoir pour effet la forme des cycles sur le signal électroglottographiques ; les fermetures demeurent plus saillantes que les « faux pics » (dûs à l'artefact de filtrage) à l'observation cycle par cycle, mais non si un seuil unique est appliqué sur toute la syllabe, car celui-ci, s'il est fixé assez bas pour que les dernières fermetures soient détectées, aboutira aussi à inclure les pics artefactuels. La même difficulté est apparue dans le traitement des données du ton montant en naxi (par le locuteur M4). Ces cas rendent illusoire la tentative de trouver une définition idéale de seuil, valable pour tous les registres phonatoires, et sur l'intégralité d'un échantillon à analyser.

- en définissant séparément les seuils pour deux moitiés de la syllabe, la frontière entre ces deux moitiés étant également déterminée par l'utilisateur au vu des signaux.

Cette solution a permis de traiter l'intégralité des signaux enregistrés au cours de la thèse, qui constituent un corpus étendu et varié (en particulier dans le domaine de la voix parlée masculine ; la difficulté de réaliser un enregistrement électroglottographique de qualité dans le cas des voix féminines a conduit d'emblée à recourir préférentiellement à des consultants masculins).

Traitement des pics de fermeture doubles ou multiples

Dans la détection des pics de fermeture, comme dans la détection des pics d'ouverture, se pose le problème des doubles pics. Le programme <peakdet> laisse à l'utilisateur la liberté de choisir le plus élevé des pics voisins, le premier, le dernier, ou une valeur intermédiaire calculée par une méthode de barycentres : celle-ci est illustrée par la figure 4.1, qui représente la courbe rééchantillonnée d'un pic de fermeture triple (données naxi, locuteur M4), phénomène relativement courant au ton montant chez ce locuteur. Une réflexion générale sur les limites de cette méthode est proposée en section 1.5.

4.1.4. Principes pour la détection des pics d'ouverture

Dans une première version des programmes de Vu-Ngoc Tuân, les pics d'ouverture étaient détectés séparément des pics de fermeture, selon une méthode similaire. Un algorithme appariait ensuite fermetures et ouvertures, de la façon suivante : si, à l'intervalle entre deux fermetures, correspondait une ouverture et une seule, le quotient ouvert était calculé ; lorsqu'aucun pic d'ouverture ne correspondait à cet intervalle, le programme affectait à ce cycle une valeur 0 de quotient ouvert (valeur de convention permettant de distinguer ces cycles des autres) ; lorsque deux pics d'ouverture (ou plus) correspondaient à l'intervalle détectés, le programme choisissait le plus élevé pour le calcul du quotient ouvert, mais affectait ce résultat d'un signe – (en faisant une valeur négative) pour signaler que le pic d'ouverture n'était pas unique.

Cette méthode ne permet pas pour autant d'exclure tous les pics incertains ou peu marqués. La saillance du pic recherché par rapport au bruit de fond du signal dérivé du signal électroglottographique est faible, surtout dans le cas de voix féminines. Lors du tri manuel des résultats, des erreurs de détection assez nombreuses sont apparues. Les critères selon lesquels ces valeurs apparaissaient comme erreurs de mesure est le degré de variation d'une période sur l'autre, et la valeur absolue : un quotient ouvert supérieur à 80 ou inférieur à 20 n'est pas plausible, sauf exceptionnellement aux marges du signal (premier cycle glottique détecté, dernier cycle détecté). Ces valeurs étaient dues au fait que certains pics d'ouverture,

tout en demeurant reconnaissables à l'œil exercé, ne se détachaient pas du bruit de fond du signal.

1.4.1. Choix de lisser le signal

La solution technique adoptée par la suite a consisté à lisser le signal électroglottographique, pour que les pics d'ouverture ressortent mieux. Une méthode consiste, pour un pas de lissage <n>, à remplacer <n> valeurs successives par leur somme, de sorte que, par exemple, la séquence

10 20 15 19 33 27 23 33 40

devienne, avec un pas de lissage de 3,

45 79 96.

La fréquence d'échantillonnage se trouve ainsi divisée par n, ce qui ôte un peu de sa précision temporelle à la mesure.

La solution retenue pour le programme <peakdet> consiste à modifier la valeur en chaque point par la moyenne des (2 *n + 1) points qui l'entourent (méthode implémentée par le script <smoo.m>), de sorte que la fréquence d'échantillonnage n'est pas modifiée, non plus que l'index de chaque échantillon, ce qui rend plus transparente la conversion des index en unités temporelles.

Le pas de lissage est choisi par l'utilisateur : cette valeur n correspond au nombre d'échantillons avant et après l'échantillon considéré qui seront moyennés. Ainsi, un pas de 0 signifie qu'aucun lissage n'est effectué, un pas de 1 signifie que le lissage prendra en compte le point précédent et le point suivant (soit trois points en tout), un pas de 2 signifie que cinq points en tout seront moyennés : deux avant, deux après, plus le point central lui-même.

Diverses expériences suggèrent qu'**un pas de lissage de 1 est adéquat pour les signaux de bonne qualité** (ceux qui apparaissent déjà lisses), et qu'un pas de lissage de 2 voire 3 améliore la détection des pics sur des signaux relativement bruités (cela pour une fréquence d'échantillonnnage de 44100 Hz). Ce n'est qu'à partir d'un pas de lissage d'environ 6 qu'une confusion peut naître entre pics voisins, qui se trouvent noyés en un pic simple, et que le lissage déforme les pics en les aplatissant. Le pas de lissage employé dans nos analyses varie de 1 (valeur adoptée comme standard) à 3 (pour les signaux des locuteurs naxi M1 et F1, enregistrés avec un électroglottographe de Fourcin, légèrement plus bruités que ceux enregistrés avec l'électroglottographe de Rothenberg, et pour la locutrice F2).

S'agissant de la méthode de détection des pics d'ouverture sur ce signal lissé, notre expérience est que la détection des minima locaux entre deux pics de fermeture donne des résultats précis et satisfaisants dans la plupart des cas.

Poussé à son extrême, le lissage se rapproche de la méthode des barycentres des régions positives et négatives décrite ci-dessous (section 1.5), expérimentée par Vu-Ngoc Tuân : sur le signal dérivé n'apparaissent plus alors que deux bosses par cycle, l'une positive, l'autre négative. Comme expliqué dans la discussion de cette méthode, elle s'éloigne de ce qui fait le grand intérêt de la méthode qui recourt à la dérivée du signal : la correspondance avérée entre les pics et des événements physiologiques. Le pas de lissage doit être choisi de façon à ne pas noyer les pics ; il existe des cas où le pic d'ouverture est double ou incertain, et il est alors de bonne logique de renoncer au calcul du quotient ouvert ; il est toujours possible de faire émerger un pic en appliquant un lissage fort, mais cela revient à créer un pic, non à faire ressortir une information présente dans le signal.

1.4.2. Détection du minimum local ou barycentre des pics : Choix visuel au vu de quatre résultats

Le programme <peakdet> effectue la détection des pics d'ouverture et le calcul du quotient ouvert selon deux méthodes : détection du minimum local entre deux fermetures, et barycentre des pics détectés. La méthode de détection du minimum local ne reflète pas la forme du signal dérivé : que le minimum local émerge à peine du signal ou soit bien visible, la détection du minimum local donnera le même résultat. En revanche, la méthode des barycentres consiste à effectuer un seuillage à 70% du minimum local, et à relever l'amplitude et la position de tous les pics qui dépassent ce seuil ; en conséquence, un pic unique et bien marqué sera seul détecté, et le barycentre sera égal au minimum local, tandis que si plusieurs pics ont une amplitude égale ou supérieure (en valeur absolue) à 70% du minimum local, leur barycentre donnera une valeur différente de la simple méthode de détection du minimum.

Le principe est de permettre à l'utilisateur de se prononcer sur la possibilité ou non de calculer les valeurs de quotient ouvert pour l'intervalle considéré, sur la base de l'inspection visuelle du signal électroglottographique et de son signal dérivé, et du degré de correspondance entre les résultats donnés par les deux méthodes : les pics multiples sont le signe d'une forme de signal non idéale ; une médiocre correspondance entre détection des minima et méthode du barycentre signale que les minima se détachent peu des pics environnants.

Enfin, chacune de ces deux méthodes est appliquée deux fois, une fois sur le signal non lissé et une fois sur le signal lissé, afin que l'utilisateur puisse évaluer l'effet du lissage et le cas échéant corriger le pas de lissage choisi. Ce sont donc en tout quatre courbes qui sont affichées pour chaque item traité.

La figure 4.2 illustre sur une syllabe les résultats de quotient ouvert qui sont calculés par ces quatre procédures. Le premier constat, dans cet exemple, est que la courbe de quotient ouvert est raisonnablement continue. Elle part d'une valeur très élevée (supérieure à 80%), ce qui est fréquent, en particulier après une consonne aspirée ; elle descend ensuite jusqu'aux environs de 55%. En revanche, la dernière valeur, subitement plus élevée, n'est pas plausible, et appelle vérification (et très certainement exclusion).

La ligne bleue, correspondant aux valeurs calculées par détection des minima sur le signal lissé, coïncide presque entièrement avec les astérisques vertes, qui correspond aux valeurs calculées par la même méthode sur le signal non lissé.

En revanche, une différence de l'ordre de 2% s'observe ici de façon assez régulière entre les méthodes par détection des minima et les méthodes de barycentre des pics, ces dernières donnant des valeurs de quotient ouvert plus basses. L'explication du phénomène se trouve dans la forme du signal dérivé, présenté et commenté sur la figure 4.3, qui montre une vue très rapprochée d'un des pics d'ouverture de ce signal dérivé : une denture apparaît, le pic étant suivi d'un second pic de plus faible amplitude. L'analyse de la forme des pics décèle un changement de signe de la dérivée seconde, ce qui signale un pic double ; une valeur intermédiaire entre les deux pics est alors calculée, prenant en compte la hauteur respective de chacun. Le pic résultant est plus proche de la fermeture suivante, d'où un quotient ouvert légèrement plus bas.

Sur la base de ces observations, l'utilisateur peut décider

- de retenir les valeurs calculées par détection des minima, jugeant qu'elles donnent un résultat cohérent et que la saillance du pic principal est suffisante pour justifier qu'on le prenne comme indication suffisante de la position de l'ouverture glottale
- de retenir les valeurs calculées par barycentre, pour refléter la forme non unique des pics
- ou de ne retenir aucune valeur de quotient ouvert pour la syllabe, jugeant que le caractère non unique du pic n'autorise pas la détermination d'un point unique d'ouverture de la glotte sur lequel fonder une mesure de quotient ouvert.

Le choix effectué dépend de l'objectif que se fixe l'étude, et du degré de finesse nécessaire pour la démonstration des hypothèses soumises à vérification expérimentale. Si l'objectif était de montrer l'absence de constriction glottale sur le type de rime correspondant à cet intervalle de signal électroglottographique, la variation de l'ordre de 2% entre méthodes de mesure serait sans importance, toutes les mesures coïncidant pour indiquer une qualité de voix moyenne (cette observation étant bien sûr à affiner par rapport à des valeurs de référence calculées pour ce locuteur) et non une tendance à la constriction glottale. Il serait alors préjudiciable à la validation statistique des observations d'exclure toutes les valeurs pour cet intervalle à cause d'une forme non idéale du pic d'ouverture. Si en revanche

l'objectif est d'étudier les différences de quotient ouvert qui peuvent exister entre voyelles, différences qui sont faibles, une indétermination de l'ordre de 2% peut ne plus être acceptable.

Le choix fait dépendra aussi de la qualité globale de l'enregistrement électroglottographique : si dans un assez grand nombre de cas l'accord est très bon entre les méthodes de calcul, il paraît raisonnable d'isoler les cas douteux, et de les traiter à part ou de les exclure tout-à-fait.

Cette étape du travail est donc d'une grande importance. Il importe d'expliciter par des exemples notre méthode dans le choix des valeurs de quotient ouvert retenues, pour indiquer comment ont été obtenues les valeurs présentées dans le présent travail.

Un cas dans lequel le calcul du quotient ouvert est particulièrement problématique est présenté dans la figure 4.4. La dispersion des valeurs obtenues par la méthode des barycentres est le signe de pics d'ouverture imprécis ou multiples. Aucune des méthodes ne donne de résultats de quotient ouvert qui forment une ligne continue. Au vu de ces résultats, il est impératif d'exclure toutes les valeurs de quotient ouvert sur la deuxième partir de la rime. Sur la première partie de la rime, les résultats obtenus par détection des minima sur le signal lissé (en bleu) ne sont pas absurdes, et sont dans une certaine mesure continus ; avant d'exclure les résultats de toute la syllabe, il est donc utile de regarder l'allure du signal dérivé du signal électroglottographique. Celui-ci est présenté sur la figure 4.5, qui correspond à l'une des premières périodes. Son allure est fortement dentelée ; les dentelures sont décelées comme autant de pics, déplaçant, d'une façon très variable d'une période sur l'autre, le centre de gravité des pics calculé par la méthode du barycentre. En revanche, il apparaît qu'il n'y a pas clairement de dédoublement de pic d'ouverture (présence de deux pics d'amplitude comparable). L'idée de retenir les valeurs de quotient ouvert calculées par détection des minima (sur le signal lissé) pour la première partie de la syllabe est donc défendable ; l'intervalle sur lequel cette mesure peut être retenue couvre les périodes 3 à 14. Les deux premières périodes, et toutes celles à partir de la période 15, sont à exclure, aucun pic ne ressortant nettement. (Le choix fait lors du traitement des données a consisté à exclure les valeurs de quotient ouvert pour toute la syllabe, au vu de la variabilité générale des résultats sur la syllabe.)

La figure 4.6 représente une situation fréquente : l'accord entre méthodes est globalement bon ; les résultats calculés d'après le signal non lissé comportent quelques valeurs visiblement absurdes, corrigées dans les mesures sur le signal lissé, ce qui indique que la qualité du signal électroglottographique n'est pas idéale. L'examen de ce signal (figure 4.7) montre que la détection des pics d'ouverture n'est pas problématique, même si le détail de la courbe est peu régulier : un léger bruit est présent dans le signal, qui pourrait être réduit en

adoptant un pas de lissage plus élevé. Dans cet exemple, les valeurs de quotient ouvert calculées par la méthode du minimum local sur le signal lissé (en bleu) ont été retenues, excluant seulement la valeur sur la 9e période, artificiellement plus élevée que les valeurs voisines.

Un exemple dans lequel la méthode du barycentre des pics permet d'obtenir une courbe de quotient ouvert continue malgré un dédoublement des pics est présenté en figure 4.8. Dans cet exemple, la forme en dents de scie de la courbe bleue (valeurs calculées en choisissant le minimum local) interdit de retenir ces valeurs. Les valeurs obtenues par barycentre des pics d'ouverture (carrés noirs) montrent un peu plus de continuité. L'examen du signal dérivé (figure 4.9) confirme qu'il s'agit d'un dédoublement de pics ; si l'on accepte le compromis que représente la méthode par barycentre des pics, celle-ci apparaît ici comme plus adaptée que la méthode par détection du minimum.

Pour finir, la figure 4.10 présente le résultat d'une analyse pleinement satisfaisante selon les critères employés dans notre étude : le résultat est presque parfaitement continu, et les résultats des différentes méthodes sont très proches. La forme des pics d'ouverture correspondants sur le signal dérivé du signal électroglottographique, présentée en figure 4.11, montre des pics uniques et bien marqués. Ce cas est le plus fréquent.

4.1.5. Réflexion sur les méthodes de détection des pics de fermeture pour le calcul de la fréquence fondamentale. Proposition de deux algorithmes.

Dans la première version du programme, suite à la détection des pics de fermeture, l'inverse de la durée séparant les pics successifs était calculé ; cette valeur était ensuite comparée à un maximum et un minimum indiqués par l'utilisateur (dans un fichier texte séparé), par exemple un minimum de 20 Hz et un maximum de 500 Hz. Les valeurs qui ne satisfaisaient pas à ces critères étaient supprimées du fichier de résultats.

L'inconvénient est que les courts intervalles de temps entre pics dédoublés sont tout simplement exclus des calculs, de sorte que la somme des durées des cycles retenus est inférieure à la durée totale de la rime (de la première fermeture à la dernière fermeture). Il paraissait donc utile d'exclure ces pics avant l'étape du calcul de la fréquence fondamentale.

La présence de doubles pics (évoquée ci-dessus, section 1.2.3) soulève des problèmes dans le calcul des paramètres (fréquence fondamentale et quotient ouvert). En effet, le pic le plus élevé peut être le premier aussi bien que le dernier ; des transitions sont observées au cours d'une syllabe, un pic apparaissant puis montant en amplitude d'un cycle à l'autre pour finalement devenir le pic le plus élevé. Des observations détaillées sur ce sujet ont été effectuées sur la base d'images de la glotte par cinématographie ultra-rapide combinées avec l'électroglottographie par plusieurs auteurs (voir en particulier Henrich, Gendrot *et al.* 2004, « Characterization of features observed on the derivative of EGG signal by the use of high

speed cinematography ») ; ces observations confirment que l'exclusion de l'un des pics pour le calcul de paramètres n'est pas un problème trivial, chacun des pics correspondant à une partie du phénomène physiologique de vibration. Un exemple est présenté sur la figure 4.12.

Une solution qui donne un résultat convenable pour notre propos consiste à retenir le pic le plus élevé. Pour cela a été réalisé un programme <picsimple>, qui s'applique entre la détection des pics de fermeture et le calcul de la fréquence fondamentale et du quotient ouvert. Son fonctionnement est le suivant : si la distance temporelle entre un pic n et le pic n - 1 est inférieure à un cinquième de la période précédente, l'algorithme écarte l'un de ces deux pics : celui qui a l'amplitude la plus faible.

Un léger saut de fréquence fondamentale s'observe par conséquent sur la courbe dans les cas où le premier pic est plus élevé sur la première partie de la rime et le second plus élevé sur la seconde partie, ou inversement. Cette situation est illustrée par la figure 4.13 : le troisième cycle glottique (dont la durée est indiquée par le segment en rouge sur la figure) sera calculé en retenant les pics les plus élevés, dont l'un est le premier, et l'autre le deuxième. Le cycle sera donc plus long, et sa fréquence fondamentale plus basse, que pour les cycles voisins, de sorte qu'on obtiendra une irrégularité dans la courbe de fréquence fondamentale, alors que l'examen de la dérivée du signal électroglottographique montre que le changement est graduel.

Dans le programme <peakdet>, l'utilisateur a le choix entre calcul de la fréquence fondamentale sur la base du pic le plus élevé ou d'un barycentre de pics : point intermédiaire entre les divers pics, d'autant plus proche d'un pic que celui-ci sera élevé. (Par souci d'exhaustivité, le choix est également proposé de sélectionner à chaque fois le premier pic, ou le dernier, dans un groupe de pics, mais après expérience ces solutions ne nous paraissent pas réellement défendables). La méthode du barycentre reflète l'intuition du caractère graduel du changement d'amplitude des pics au cours de la syllabe ; il a pour effet de lisser la transition lorsqu'un pic apparaît et grandit en amplitude au point de devenir le plus élevé.

La méthode des barycentres de pics ne fait néanmoins que repousser le problème à un autre niveau : celui du seuil fixé pour la détection des pics de fermeture. En effet, cette méthode permet d'éviter les décrochements dans la courbe de fréquence fondamentale correspondant à des items tels que celui représenté dans la figure 4.13, mais peut aboutir à des décrochements similaires dans des cas où le pic secondaire est proche de la valeur-seuil choisie pour la détection des pics. Si deux périodes successives ont des pics de fermeture avec un pic secondaire de faible amplitude, que ce pic secondaire soit juste au-dessous du seuil dans le premier cas et juste au-dessus dans le second, le premier pic sera considéré comme un pic unique et le second comme un double pic, pour lequel la méthode des

barycentres de pics s'appliquera, avec pour effet une irrégularité plus ou moins importante dans la ligne de fréquence fondamentale.

Une proposition plus radicale a été formulée par Vu-Ngoc Tuân pour sortir de ces difficultés. Il s'agit de ne plus raisonner en termes de pics, mais en termes de partie positive et partie négative du signal, la position de la fermeture et de l'ouverture étant définie en termes de barycentre de ces régions dans leur globalité. L'inspiration de cette méthode vient de l'expérience des pics d'ouverture peu marqués, difficiles et parfois même impossibles à détecter ; cette méthode permet de tenir compte de toute la région d'ouverture, et de calculer, dans tous les cas sans exception (quelle que soit la forme du signal dérivé du signal électroglottographique), une valeur correspondant en un certain sens au centre de cette région : son barycentre.

Il y a quelque imprudence à s'orienter dans ce sens : l'intérêt de la méthode employant la dérivée du signal électroglottographique, telle qu'elle est décrite par Henrich, d'Alessandro *et al.* 2004, tient pour l'essentiel à la correspondance (établie par de nombreuses expériences) entre les événements que l'on relève sur la dérivée du signal électroglottographique et des événements physiologiques ayant une influence sur le signal acoustique. Le dédoublement des pics pose certains problèmes, mais il paraît préférable de les traiter sans quitter entièrement le cadre qu'ils offrent. La méthode du barycentre des régions positives et négatives s'éloigne de ce terrain sûr ; elle n'a pas été retenue dans le présent travail.

Les deux méthodes de sélection de pic pour le calcul de la fréquence fondamentale qui ont été utilisées dans les analyses rapportées ici sont donc : sélection du pic le plus élevé lorsqu'il n'a pas été constaté (par inspection visuelle) de doubles pics de fermeture fréquents ; méthode des barycentres de pics lorsque les doubles pics de fermeture sont fréquents. Les résultats sont proches dans les deux cas.

Conclusion au sujet de la fréquence fondamentale

Il importe de signaler que **la mesure effectuée représente une estimation de la durée des cycles glottiques plutôt qu'une mesure de fréquence fondamentale, l'emploi de la notion de fréquence fondamentale supposant une périodicité, tandis que nos mesures ne posent à aucun moment de critère de périodicité**. Cette approximation paraît en adéquation avec les données : ce qui fait l'objet de l'étude, ce sont les cycles glottiques au cours de certaines rimes, dans des langues données ; le critère de périodicité des cycles glottiques n'a pas en parole l'importance qu'il a en voix chantée (type de production pour lequel a été proposé l'algorithme par corrélation DECOM, *Derivative-Electroglottographic Correlation-based Method* : Henrich, d'Alessandro *et al.* 2004). C'est pourquoi il n'a pas paru utile au cours de l'étude d'enrichir le programme <peakdet> d'un module de détection de la fréquence

fondamentale par corrélation, qui s'appliquerait sur la partie quasi-périodique du signal, seules les franges étant traitées par détection directe des pics. Les quelques problèmes qui demeurent pour une estimation très précise de la fréquence fondamentale suggèrent qu'une intégration de DECOM et de la méthode par détection des pics est souhaitable, DECOM s'appliquant en premier, et la détection des pics étant ensuite appliquée avant la première période de voisement relevée par DECOM et après la dernière période, permettant d'étendre les mesures aux premiers et derniers cycles glottiques, moins réguliers. L'implémentation logicielle de cette méthode combinée pourra être réalisée pour le propos d'études ultérieures.

4.2. Interface de vérification de valeurs de formants

Les bornes de début et de fin pour la détection de formants sont fondées sur les résultats de l'analyse électroglottographique, ce qui permet que la détection coïncide précisément avec l'intervalle de voisement correspondant à la voyelle : du premier au dernier pic de fermeture détectés sur la dérivée du signal électroglottographique pour la rime considérée. Le cahier des charges du script MATLAB baptisé <forver> (pour FORmant VERification) est le suivant :

- calculer les intervalles temporels sur lesquels l'analyse doit être réalisée, à partir du fichier de résultats de l'analyse électroglottographique
- pour chaque voyelle à traiter : lancer les algorithmes de détection de formants (Praat et ForEst) ; afficher formant par formant les estimations de valeurs formantiques (de Praat ou de ForEst, au choix de l'utilisateur), et demander à l'utilisateur de vérifier, lui permettant de supprimer des valeurs (techniquement, celles-ci sont fixées à zéro, et exclues des calculs ultérieurs).

L'intervalle de signal audio est affiché et joué, ce qui permet une vérification supplémentaire de la position des bornes de début et fin.

4.3. Implémentation d'un test-t apparié

L'environnement de programmation MATLAB propose uniquement un test-t non apparié, et ne renvoie pas la valeur de t, seulement la valeur de p. Un script a donc été créé, qui implémente le test-t apparié, d'après les formules données dans le *Manuel d'Analyse des Données sur Ordinateur : application aux sciences du langage et de la parole* de J.-Y. Dommergues (Laboratoire de Phonétique de l'Université Paris 7 et Laboratoire de Phonétique et Phonologie de l'Université Paris 3, années 1999-2002).

4.4. Simulations syllabiques

Le script <aver> est conçu pour calculer des courbes moyennes de fréquence fondamentale et quotient ouvert à partir d'ensembles de syllabes de longueurs différentes, et pour lesquelles il manque certaines valeurs de quotient ouvert (dans les cas où la valeur de quotient ouvert ne peut être calculée pour telle ou telle période).

4.4.1. Explication au sujet du nombre de points des courbes rééchantillonnées

L'algorithme interpole la courbe de fréquence fondamentale et de quotient ouvert de chacune des syllabes données en entrée, ce qui aboutit à un vecteur en un nombre de points donné par la variable <samplenumber> points. Selon les cas, une courbe en 3 ou 5 points peut être considérée comme suffisante par l'utilisateur. Il paraît néanmoins prudent de choisir une valeur élevée au stade du moyennage des valeurs ; l'inspection visuelle de la courbe rééchantillonnée en 100 points permet ensuite de se prononcer sur l'adéquation d'une simplification en un plus petit nombre de points (10, 5, 3 ou même un seul). La valeur par défaut de <samplenumber> est donc fixée à 100 (sur le conseil de J.-Y. Dommergues, communication personnelle), ce qui ne pose aucune difficulté technique au vu des performances actuelles des ordinateurs.

Il importe de souligner que **le choix de rééchantillonner toutes les courbes en 100 points n'implique nullement qu'une valeur en un de ces 100 points ait une quelconque valeur par elle-même**, puisque dans cette division en 100 points, chaque point correspond à moins d'un cycle glottique. Le choix d'employer 100 points, plutôt que 20, 10 ou même 3 tient au souhait de ne pas dégrader la précision des mesures en amont du résultat final, de même que dans les sciences physiques les résultats numériques ne sont arrondis qu'à l'étape du résultat et non à chaque étape du calcul, ce qui introduirait en fin de calcul une incertitude élevée. Pour prendre un exemple tout théorique : soit l'opération suivante :

$[\pi * 5.3491 * 24.44842]^2 = 169349.0396$

Au terme du calcul, le résultat peut être arrondi en fonction de l'incertitude expérimentale des données fournies en entrée, et de la précision désirée. Si en revanche chaque élément est arrondi à une décimale avant l'opération, cela introduit une incertitude multipliée, au point que le résultat devient différent dès le deuxième chiffre significatif (1.7×10^5 dans le premier cas, 1.6×10^5 dans le cas où les valeurs ont été arrondies avant le calcul) :

$[3.1 * 5.3 * 24.4]^2 = 160714.3957$

4.4.2. Traitement des cas où certains points de données manquent

Lorsque certains points de données de quotient ouvert manquent dans le fichier donné en entrée, les points qui correspondent à l'intervalle de temps concerné sont pareillement laissés non renseignés. Cette procédure est préférée à une interpolation des points manquants : ainsi, une rime de 200 ms qui compterait 20 cycles glottiques dont seuls les

cinq premiers et les trois derniers présenteraient des pics d'ouverture uniques et bien marqués sur la dérivée du signal électroglottographique pourrait, sans difficulté technique, voir sa courbe de quotient ouvert interpolée en fonction de ces valeurs ; mais cela reviendrait à introduire des points de données, ce qui ne paraît nullement légitime.

4.4.3. Des moyennes aux simulations syllabiques

A chaque rime syllabique analysée correspond une ligne dans une matrice de <samplenumber> colonnes. Chacun des paramètres donnés en sortie par le programme <peakdet> est ainsi interpolé : fréquence fondamentale, quotient ouvert, DECPA, amplitude des pics d'ouverture sur le signal dérivé du signal électroglottographique. La moyenne de toutes les lignes donne un vecteur moyen en <samplenumber> points. L'écart-type est également calculé, permettant de comparer la variabilité en chacun des points.

La durée de la rime, définie comme la durée qui sépare la première fermeture glottique de la dernière fermeture glottique, est moyennée sur l'ensemble des syllabes concernées (par exemple, pour les données anglaises, les 24 rimes de syllabe accentuée de disyllabe, ensemble distinct des 24 rimes de syllabe non accentuée de disyllabe, et des 24 rimes de monosyllabes). Les courbes moyennes peuvent donc être représentées en fonction de la durée.

La durée qui sépare chacun des points de la courbe rééchantillonnée est la même, ce qui est commode pour l'analyse statistique, mais s'éloigne en revanche de la division en cycles glottiques qui caractérise les données fournies en entrée. Il est souhaitable de disposer d'une courbe moyenne sur laquelle il soit aisé de relever l'évolution d'un cycle glottique à l'autre, par exemple pour visualiser la différence moyenne de fréquence fondamentale et de quotient ouvert qui sépare les trois derniers cycles glottiques de tel ou tel ensemble de syllabes. Une seconde partie du programme <aver.m> est consacrée à cette tâche : la création d'une *simulation syllabique*, sorte de *type idéal* de la syllabe.

Un vecteur-temps est créé pour établir une correspondance entre chacun des points et la longueur moyenne des syllabes ; <LEN> est le vecteur contenant la longueur de chacune des syllabes :

```
meanlength = mean(LEN);

timestep = meanlength/(samplenumber-1);

timevector(1) = 0;

for i = 2:samplenumber

   timevector(i) = timevector(i-1)+timestep;
```

end

Le premier cycle de la courbe simulée est le premier point de données de la courbe en <samplenumber> points. La durée de ce cycle est l'inverse de la valeur de fréquence fondamentale : par exemple, pour une fréquence fondamentale de 150, la durée du cycle est de 6.66 ms. Les paramètres du second cycle sont donc obtenus en interpolant une valeur à 6.66 ms sur la courbe moyenne. Jusqu'à ce que le programme parvienne à la fin de la syllabe, la séquence d'opérations effectuées est la suivante, où <ATV>, Average Time Vector, est le vecteur-temps de la courbe simulée, et <AFV>, Average Fundamental frequency Vector, contient les valeurs de F_0 correspondantes :

ATV (i) = ATV (i-1) + (1/AFV(i-1));

AFV (i) = interp1(timevector,meanFovector,ATV(i));

c'est-à-dire que la position du point temporel suivant est calculé en ajoutant la durée de la période à la valeur précédente ; la valeur de F_0 en ce point est calculée en interpolant d'après la courbe moyenne (en <samplenumber> points) rapportée au vecteur-temps <timevector> calculé par les lignes exposées plus haut.

Ce mode de calcul, qui part du début de la syllabe et progresse de cycle en cycle, ne permet pas une correspondance exacte entre longueur de la courbe simulée et longueur moyenne des rimes mesurées : parvenu en fin de rime, il n'est pas indiqué d'interpoler une valeur de fréquence fondamentale au-delà de la durée moyenne constatée ; inversement, interrompre le calcul au dernier cycle glottique qui demeure en-deçà de la durée moyenne de la rime, c'est exclure les valeurs finales, qui sont parfois cruciales. Le programme <aver.m> calcule donc un coefficient de correction :

correctioncoefficient = meanlength / ATV(PN);

(rappel : <meanlength> est la longueur moyenne des syllabes mesurées, <ATV(PN)> est le dernier point temporel auquel le programme aboutit lors de sa construction période par période du vecteur-temps de la rime simulée, <ATV>)

Le vecteur <ATV> est ensuite multiplié par ce coefficient, et chacune des valeurs de F_0 et quotient ouvert de la syllabe simulée est recalculée en fondant l'interpolation sur le nouveau vecteur ATV. De la sorte, le dernier point de la simulation correspond au dernier point de la courbe rééchantillonnée en <samplenumber> points.

4.4.4. Perspectives de développement futur

L'approximation par une courbe mathématique *(curve fitting)* n'a pas été réalisée pour les données présentées ici ; cette méthode, notamment employée par Andruski et Costello 2004, « Using polynomial equations to model pitch contour shape in lexical tones: an example from Green Mong » et Vu-Ngoc Tuân, d'Alessandro *et al.* 2005, pourra à l'avenir être intégrée au programme <aver>.

Table des matières de l'Annexe 3 (« Perspectives théoriques complémentaires »)

1. L'alignement des points-clefs de la courbe de fréquence fondamentale avec la « ligne des segments » : la problématique et sa remise en cause _____ 423

1.1. Le déplacement du pic de F_0 correspondant à l'« accent de phrase » a un effet sur l'interprétation de l'énoncé _____ 424

1.2. Limites de la problématique : l'alignement est-il un paramètre indépendant des autres paramètres phonétiques ? _____ 426

1.2.1. La fréquence fondamentale n'est pas un paramètre indépendant des autres paramètres phonétiques _____ 427

1.2.1.1. Existence de relations d'équivalence et de compensation entre paramètres _____ 427

1.2.1.2. Des données expérimentales établissent le rôle de l'intensité et de la pente de F_0 dans la perception des contrastes d'alignement du pic de F_0 _____ 427

1.2.2. Le mouvement des articulateurs, comme la courbe de F_0, est continu _____ 428

2. Tons et structure métrique en chinois _____ 430

Annexe 3. Perspectives théoriques complémentaires

1. L'alignement des points-clefs de la courbe de fréquence fondamentale avec la « ligne des segments » : la problématique et sa remise en cause

La problématique qui a d'abord été envisagée pour la présente thèse concernait l'alignement des points-clefs de la courbe de fréquence fondamentale avec les segments et les structures syllabiques. La réflexion menée ci-dessous au sujet de cette problématique conclut à la nécessité d'une prise en compte de la multiplicité des paramètres prosodiques.

L'attention prêtée, dans le courant autosegmental-métrique, aux maxima et minima de fréquence fondamentale, considérés comme indices de tons binaires H et L, croise une autre problématique de recherches : l'étude de l'alignement des courbes de fréquence fondamentale avec les structures syllabiques et les segments. L'idée centrale est que la façon dont les courbes mélodiques (niveau *glottique*) sont coordonnées avec l'articulation (niveau *supraglottique*, celui des « segments ») est contrôlée à des fins linguistiques.

L'exploration de ce domaine a commencé avec les premiers outils informatiques permettant de modifier artificiellement la fréquence fondamentale (ci-après F_0) d'un signal de parole ('t Hart et Cohen 1973, « Intonation by rule: a perceptual quest »). Gösta Bruce observe qu'en suédois, la position d'un *pic* ou d'une *vallée* de F_0 dans l'énoncé (« reaching a certain pitch level at a particular point in time (...), not the movement (rise or fall) itself ») est plus régulière que la forme globale de la courbe (Bruce 1977:132). Cette question a été approfondie par de nombreux auteurs, dont les orientations théoriques sont variées[1]. Au-delà de l'étude de l'*intonation*, domaine dans lequel cette problématique est d'abord apparue, le champ de recherche a récemment été étendu aux *tons lexicaux* du chinois mandarin (Xu Yi 1999) et d'une langue à tons d'Afrique, le kinyarwanda (Myers 2003)[2]. **Diverses**

[1] Voir en particulier Kohler 1987, 1991c ; House 1990, 2003b ; Pierrehumbert et Steele 1989 ; Silverman et Pierrehumbert 1990 ; Arvaniti, Ladd *et al.* 1998 ; Ladd, Faulkner *et al.* 1999 ; Ladd, Mennen *et al.* 2000 ; d'Imperio 2000, ainsi que les documents mis à disposition par Jill House : http://www.phon.ucl.ac.uk/home/shl9/jill/house.htm.

[2] Cette extension est liée au postulat des théories autosegmentales-métriques selon lequel l'intonation serait à modéliser en termes de « tons », postulat qui amène à considérer qu'un pic de fréquence fondamentale peut être l'indice d'un « ton Haut » dans une langue germanique ou romane sans tons lexicaux aussi bien que dans une langue qui possède des tons lexicaux : voir en particulier Ladd 1996, et la Discussion proposée au chapitre IV, section 2.

considérations, dont les résultats obtenus au laboratoire de phonétique de Kiel par la prise en compte de l'intensité et de la pente mélodique, nous paraissent en fait remettre en cause le statut de variable indépendante que beaucoup de ces études accordent à la position temporelle des pics de F_0.

1.1. Le déplacement du pic de F_0 correspondant à l'« accent de phrase » a un effet sur l'interprétation de l'énoncé

Une étude qui porte sur l'allemand conclut au **caractère significatif, au plan pragmatique, de la position du pic de F_0 qui correspond à l'accent de phrase**[1] (Kohler 1987, 1991c « Terminal intonation patterns in single accent utterances of German : phonetics, phonology, and semantics »). L'expérience consiste à varier artificiellement la position du pic dans la phrase suivante :

Sie	hat	ja	gelogen.
3e sg. féminin	verbe auxiliaire « avoir »	particule	mentir (participe passé)

« Elle a menti. »

Dans cette phrase, l'unique accent lexical se place sur la seconde syllabe du verbe : *gelogen*. Cette même syllabe se trouve également porteuse de l'accent de phrase[2]. Après enregistrement de cette phrase (prenant soin de la réaliser avec une intonation relativement neutre), la modification de sa courbe de F_0 a lieu de façon systématique et contrôlée (« ...systematic and controlled parameter manipulation in computer signal processing, on the basis of hypotheses derived from preliminary production data analysis » : Kohler 1991a:14).

Le résultat de tests de perception sur les stimuli créés en déplaçant le pic est le suivant : « Early F_0 peaks signal established facts, middle peaks new information, and late peaks emphasis and contrast » (Kohler 1991b:307). Dans ce modèle ternaire, l'opposition *précoce/moyen* et l'opposition *moyen/tardif* ne sont pas symétriques : à la première correspond, au plan sémantico-pragmatique aussi bien qu'au plan perceptif, une différence catégorielle ; à la seconde, une différence scalaire de degré d'insistance[3].

> If an F_0 peak in a single-accent terminal utterance is shifted left from a medial position in the stressed syllable nucleus a categorical change occurs in perception,

---

[1] Au sujet de cette notion (en anglais *sentence stress*), nous renvoyons aux publications de K. Kohler : Kohler 1987, 1991c. Voir également les références fournies dans l'entrée correspondante du Glossaire.

[2] Voir note précédente.

[3] « The phonetic opposition signals the semantic "closed-open" or "established-new" dichotomy in the speaker's assessment of the communicative situation or of the speech acts in it, but in the late rise the openness connotation is gradually attenuated as the initial low F_0 is extended, introducing the speaker's views in opposition to the world around him. » (Kohler 1991c:145)

> which is correlated with a semantic change along the dimension "new-established". The corresponding realignment to the right produces a gradual auditory change correlated with a semantic continuum expressing degrees of emphasis. (Kohler 1991a:23)[1]

Ces propositions concernent l'allemand ; au-delà de cette langue, des tests de perception menés (dans la même étude) avec des non germanophones tendent à établir le caractère universel de la perception catégorielle[2] du caractère précoce ou médian d'un pic de F_0 : ainsi, des consultants locuteurs du chinois mandarin sont sensibles au basculement entre pic *précoce* et *médian*, qu'ils réinterprètent dans les termes des catégories tonales lexicales de leur langue maternelle. La validité de ces observations au-delà du domaine de l'allemand ressort d'autres études expérimentales suivant un protocole similaire : par exemple, Johan 't Hart, étudiant l'intonation du néerlandais, rapporte que dans cette langue deux variantes d'une même phrase peuvent se distinguer par le caractère plus ou moins tardif du pic de F_0 sur la syllabe accentuée. Commentant une phrase (qui signifie *Peter n'est jamais en retard*) avant et après déplacement de ce pic vers la droite, il écrit :

> Whereas (a) will be interpreted as a straightforward establishment of a fact, (b) has the implication (for this sentence) that the speaker is anxious that something terrible may have happened to Peter, since usually, he is never late. ('t Hart 1998:97)

Remarquons que 't Hart ne retient ici que l'opposition catégorielle entre deux types d'alignement (ce que Kohler décrit comme l'opposition entre pic précoce et pic médian), sans évoquer le troisième : l'augmentation progressive d'insistance perçue par les auditeurs à mesure que le pic est retardé (autrement dit, déplacé vers la droite) à partir de la position médiane. Le caractère non catégoriel de ce dernier phénomène en fait peut-être un objet d'étude moins attrayant pour les chercheurs : opérer avec des catégories discrètes est un objectif central en linguistique. Mais **ne retenir qu'une partie des résultats fait peser le risque**

---

[1] S'agissant de l'opposition catégorielle entre pic précoce et pic médian, l'écoute des stimuli de l'expérience menée par Klaus Kohler nous suggère que **la différence pourrait également se décrire en termes d'*association à telle ou telle syllabe***, plutôt que d'*alignement temporel précoce ou médian avec la syllabe lexicalement accentuée* : le pic *précoce* reviendrait en fait à un déplacement du pic vers la première syllabe, le *ge* de *gelogen* ; sa différence d'avec le pic *médian* serait alors une différence d'ancrage syllabique : sur la première syllabe dans un cas (notation informelle : gelogen), sur la deuxième dans l'autre (gelogen). Cette conjecture revient à suggérer que le basculement catégoriel décelé dans l'expérience en question pourrait correspondre à un changement d'affiliation syllabique du pic, plutôt qu'à une différence temporelle dans l'alignement avec une même syllabe.

[2] L'emploi du terme « perception linguistique catégorielle » est autorisé par la réalisation de *tests de discrimination* joints aux *tests d'identification*, tels qu'ils sont préconisés dans le modèle le plus répandu de la perception catégorielle (voir par exemple Harnad 1987). Ce protocole, pour répandu qu'il soit, n'est pas sans défauts, comme le soulignent Kohler 2004 (« Categorical speech perception revisited ») et Niebuhr et Kohler 2004, qui concluent : « ... the categorical perception concept needs revision ».

d'une compréhension incomplète. C'est ce qui nous paraît s'être produit dans certaines études.

Dans le courant autosegmental-métrique, Robert Ladd accorde une place de choix aux pics de F_0, comme en témoigne son article au titre programmatique : « Phonological features of intonational peaks » (Ladd 1983b). Les « tons » postulés par cette approche, *associés* à une syllabe, seraient en outre *alignés* avec les segments de façon catégorielle (Robert Ladd parle d'« Association of text to tune » : Ladd 1992:324). Mary Beckman et Janet Pierrehumbert citent les études sur l'alignement comme une *« success story »* (Beckman et Pierrehumbert 1992). La formule est reprise par Docherty et Ladd :

> ... as Beckman and Pierrehumbert note in their commentary, the modeling of high tones (fundamental-frequency peaks) is 'one of the success stories of laboratory phonology'. (Docherty et Ladd 1992:xiv)

Cette citation fournit une occasion de constater le **statut quelque peu incertain de la notion de « ton » telle qu'elle est employée dans le courant autosegmental-métrique d'études prosodiques : entité à la fois supérieurement abstraite et extrêmement concrète, identifiée au paramètre phonétique qu'est le maximum local de fréquence fondamentale**[1].

1.2. Limites de la problématique : l'alignement est-il un paramètre indépendant des autres paramètres phonétiques ?

Il paraît nécessaire de revenir sur l'idée (présentée comme un acquis par de nombreuses études) selon laquelle l'alignement de la courbe de F_0 avec la réalisation des phonèmes serait un paramètre indépendant des autres paramètres phonétiques (tels que la durée, l'intensité, l'allure générale de la courbe de F_0, l'amplitude des pics de F_0).

[1] Pour souligner les incertitudes qui demeurent sur le fond, nous prendrons pour exemple les commentaires de Jennifer Cole sur l'article de Peng Shu-Hui intitulé « Integrating the phonetics and phonology of tone alignment », commentaires qui nous paraissent représentatifs du *courant dominant* des études sur l'alignement. Jennifer Cole fait figurer l'alignement dans l'ensemble restreint de propriétés (qu'elle nomme « contrastive properties of intonation », Cole 2000:169) qui servent de primitives dans un système formulé dans le vocabulaire de la théorie de l'optimalité (Prince et Smolensky 2004). Les traits pertinents en matière d'intonation seraient les suivants : deux niveaux (« pitch levels : H and L »), des contours : « sequences over the alphabet {H,L} », et l'alignement temporel : « the location of pitch peaks (H) and valleys (L) with respect to the following points of reference : 1. the accented syllable, 2. the edge of the word domain, 3. the edge of the intonation phrase, 4. adjacent tones. » L'alignement est donc présumé être une composante essentielle des systèmes prosodiques ; en revanche, l'auteur réserve prudemment son jugement sur la réalisation phonétique de ce paramètre (Cole 2000:168, 174, 178) : ainsi, l'auteur ne soulève pas la question de savoir si l'alignement aurait lieu avec la syllabe entière ou avec la *rime syllabique* (autrement dit, avec la transition entre consonnes initiales et noyau vocalique).

1.2.1. La fréquence fondamentale n'est pas un paramètre indépendant des autres paramètres phonétiques

1.2.1.1. Existence de relations d'équivalence et de compensation entre paramètres

L'existence de relations de compensation entre les diverses caractéristiques de la courbe de F_0, la longueur, la durée, l'intensité et d'autres paramètres remet en cause le statut de paramètre autonome prêté à la position du pic de F_0. Ces relations de compensation sont connues de longue date, au plan psychoacoustique et au plan psycholinguistique ; elles sont notamment constatées dans la parole par les études suivantes : Noteboom, Brokx *et al.* 1978, Rossi et Di Cristo 1980, Repp 1982, et Pisoni et Luce 1987.

Au plan de la fréquence fondamentale, le passage d'un pic médian à un pic tardif est perçu comme une *insistance* (voir ci-dessus, section 1.1) ; cela peut être rapporté au fait qu'un mouvement plus ample demande plus de temps, et atteint plus tardivement sa cible : il est imaginable que le fait qu'une même cible soit atteinte tardivement soit perçu comme indice d'un pic plus élevé. Cette relation d'équivalence a été vérifiée expérimentalement par David House : un pic plus élevé peut être perçu de façon similaire avec un pic tardif (House 2003b, 2003a). Jacqueline Vaissière (séminaire de doctorat) relève avec humour la différence de traitement entre la stylisation extrême de l'espace mélodique (réduit, pour l'essentiel, à deux niveaux, haut et bas) et l'attention prêtée, dans le domaine temporel, à des différences qui se mesurent en centièmes de secondes. L'importance de l'allure générale de la courbe est pourtant connue (voir les références citées dans Vaissière 2002, en particulier les travaux d'Ilse Lehiste).

Les données expérimentales évoquées ci-dessous nous paraissent également aller dans le sens d'une réintégration de la problématique de l'alignement au sein d'un ensemble plus large.

1.2.1.2. Des données expérimentales établissent le rôle de l'intensité et de la pente de F_0 dans la perception des contrastes d'alignement du pic de F_0

Une expérience récente reprend le protocole de l'expérience de Klaus Kohler citée en début de section 2.1, le complétant d'une manipulation systématique de la pente de la courbe de F_0 (montée vers le pic, et redescente après le pic), et de l'intensité acoustique (Niebuhr 2003). De cette étude, il ressort que **les changements de pente de F_0 et d'intensité modifient la perception que les auditeurs ont des énoncés en question.** Cette conclusion n'est pas surprenante au vu de l'importance reconnue des relations de compensation entre ces paramètres ; en revanche, elle confirme qu'il peut être trompeur d'étudier la position temporelle d'un pic de F_0 sans tenir compte des autres paramètres.

Enfin, **au plan expérimental, la manipulation de la seule F_0** (par exemple par les algorithmes PSOLA) **ne va pas sans soulever des problèmes de taille**, car elle crée des signaux qui ne correspondent plus nécessairement aux possibilités de l'appareil vocal humain, lequel est ainsi constitué qu'un changement de F_0 s'accompagne de changements variés, notamment d'intensité (Sorin 1981, Fant, Liljencrants *et al.* 1985) ; les résultats de ces expériences demandent donc à être interprétés avec prudence. **En conclusion, il se pourrait que certains modèles reposent sur une interprétation simplifiée des observations.**

1.2.2. Le mouvement des articulateurs, comme la courbe de F_0, est continu

L'idée d'un alignement des courbes mélodiques avec les segments soulève en outre des questions délicates dès lors que l'attention se porte sur le détail phonétique de ce qui compose ces segments. Les phonèmes forment certes une chaîne d'éléments discrets, et un spectrogramme paraît offrir des points de repère précis auxquels rapporter la courbe de F_0 ; mais la réalisation des phonèmes est loin d'être strictement linéaire : ainsi, pour ne prendre qu'un exemple, le [s] de *structure* est arrondi par anticipation du /y/. L'ampleur des effets de coarticulation, entre phonèmes voisins et à plus longue distance, est bien connue (au sujet du cas extrême de / ɹ / et / l /, voir West 2003). En outre, la question de l'ancrage de la courbe de F_0 par rapport aux segments peut se retourner : quel est le *point d'ancrage* des « segments » ? **L'articulation a un caractère continu** (Fant 1974:21 et références citées).

> All work examining the detail of speech performance has worried about the relation between the discreteness of a segmental representation, on the one hand, and, on the other, the physical speech event, which is more nearly continuous and where such discrete events as can be discerned may correspond poorly with segments. (Nolan 1992:260)

L'observation du mouvement des articulateurs fournit des courbes continues : mesure du degré d'aperture de la mâchoire, de l'aire de contact entre tel point de la langue et le palais, mesure de la hauteur de tel point de la langue. Ces courbes ont leurs pics d'amplitude (maxima et minima) et leurs pics de vélocité (maxima et minima des courbes dérivées) ; le travail du spécialiste qui s'efforce d'isoler les événements principaux dans ces données soulève des problèmes non moins délicats que l'interprétation d'une courbe mélodique[1].

Dans cette perspective, **l'étude de l'alignement entre segments et courbes mélodiques appelle une prise en compte des différences entre segments : l'alignement est fortement influencé par la composition phonémique de l'énoncé, du fait que l'articulation**

[1] Ce constat rejoint la critique qu'adresse Mario Rossi à l'opposition entre *segmental* et *suprasegmental* (Rossi 1999:20-33).

supraglottique (consonantique et vocalique) a un effet sur les événements glottiques. Il est établi que l'articulation vocalique et consonantique a une influence mesurable sur la fréquence fondamentale, phénomène de *F_0 intrinsèque et co-intrinsèque* principalement dû à la variation d'impédance du conduit vocal (la pression transglottique dépend en effet de la pression intra-orale, elle-même liée au degré de constriction) ainsi qu'à la position de la langue, qui exerce indirectement une action sur le larynx (Honda 2004). S'agissant du F_0 intrinsèque des voyelles, toutes choses égales par ailleurs, la fréquence fondamentale d'un [a] est légèrement plus basse que celle d'un [i] (différence qui, dans une articulation soignée, dépasse dix Hertz). Le dévoisement des voyelles fermées dans certaines langues, et dans certains contextes positionnels (par exemple /i/ et /u/ en japonais), illustre de façon particulièrement saillante l'influence de l'articulation vocalique sur le niveau glottique, attestée dans diverses langues (Di Cristo 1981, Ladd et Silverman 1984, Shadle 1985, Whalen et Levitt 1995). S'agissant de l'effet des consonnes (F_0 co-intrinsèque), l'abaissement de la fréquence fondamentale au cours des occlusives voisées, du fait de la difficulté croissante de maintenir le voisement à mesure qu'augmente la pression intra-orale, laisse une trace décelable dans la courbe de fréquence fondamentale de la voyelle qui suit : la fréquence fondamentale du [a] de [pa] commence à une valeur plus élevée que celle du [a] de [ba], toutes choses égales par ailleurs s'entend. La figure 0.1, construite à partir de données anglaises décrites en détail au chapitre III, illustre l'effet des consonnes sur la fréquence fondamentale de la voyelle qui suit : elle représente les courbes moyennes de F_0 (ainsi que de quotient ouvert, indication sur la qualité de voix présentée en détail au chapitre I, section 1.2) de 192 monosyllabes anglais lus en phrase-cadre ; la courbe en bleu correspond à des syllabes à initiale *lenis* (traditionnellement notées /b/, /d/ et /g/, et réalisées [p], [t] et [k] en position initiale de mot), celle en rouge à des syllabes à initiale fortis (traditionnellement notées /p/, /t/ et /k/, et réalisées [p^h], [t^h] et [k^h] en position initiale de mot). Cette figure permet d'observer visuellement le fait que les rimes qui succèdent à une initiale *lenis* partent de moins haut (différence appréciable, proche d'1 ton musical sur le premier cinquième de la syllabe), et sont plus longues (d'environ 15%) ; le quotient ouvert est plus élevé juste après les consonnes phonétiquement aspirées, conséquence du flux d'air élevé associé à l'aspiration. De tels phénomènes de modification *co-intrinsèque* de F_0 se trouvent phonologisés dans certaines langues (voir en particulier Hyman et Mathangwane 1998, « Tonal domains and depressor consonants in Ikalanga »), et ont abouti à la transphonologisation de l'opposition entre occlusives voisées et non voisées en opposition tonale dans d'autres langues (*via* un stade où la distinction est portée par une opposition de qualité de voix ; le phénomène évoqué plus en détail dans la Discussion, chapitre IV).

La problématique de l'*alignement* ainsi élargie paraît représenter un programme de recherche stimulant, proche de celui de la « prosodie articulatoire »[1] ; pour être mené à bien, ce programme nous paraît demander une ouverture de champ expérimentale supérieure à celle des études sur l'alignement citées en tête de la présente section : l'organisation temporelle des segments n'est pas une donnée simple. Si l'organisation gestuelle d'une occlusive bilabiale diffère de celle d'une occlusive vélaire, la façon dont ces deux ensembles de gestes seront synchronisés avec les gestes laryngiens comporte vraisemblablement des différences.

Telle est en effet l'orientation que prennent actuellement les travaux des spécialistes, qui après avoir isolé, dans les courbes de F_0, les points considérés comme cruciaux, étudient leur alignement avec les courbes correspondant aux mouvements des articulateurs (par exemple d'Imperio, Nguyen *et al.* 2003)[2] ; le choix des points-clefs sur ces dernières (tels que : maxima des trajectoires articulatoires, ou pics de vélocité de l'articulateur primaire) soulève des problèmes similaires à ceux rencontrés dans la stylisation des courbes de F_0, cela d'autant plus que le petit nombre de capteurs employés en articulographie électromagnétique (EMA) ne permet pas encore une simulation du conduit vocal dans sa totalité, ce qui serait nécessaire pour apprécier les conséquences acoustiques des mouvements observés.

Pour ces raisons, l'alignement des courbes mélodiques n'a pas été placé au centre de la présente recherche. Les paragraphes qui suivent exposent les questionnements qui ont remplacé ce premier projet, partant des perspectives les plus larges pour aboutir à la problématique finalement retenue.

2. Tons et structure métrique en chinois

Un rapide aperçu des travaux sur la « structure métrique » du chinois vise à offrir une perspective complémentaire sur la variété des niveaux de structuration prosodique des langues à tons.

Dans sa thèse de doctorat d'Etat (*L'Accentuation en pékinois*), Monique Hoa emploie l'expression d'« accent lexical » au sujet du chinois mandarin de Pékin. La discussion

[1] Voir Vaissière 1989:29-30, Beckman et Edwards 1992, Fougeron 1999, Lœvenbruck 2000, Keating, Cho *et al.* 2003, Tabain 2003b, 2003a, Tabain et Perrier 2005, Cho 2005. L'intérêt pour l'influence du contexte prosodique sur la réalisation des segments remonte au-delà, en particulier aux travaux de Georges Straka ; bien avant l'étude expérimentale, les effets diachroniques de cette influence avaient été constatés par la phonétique historique (comme il est rappelé ci-dessous, section 3.4).

[2] Dans les travaux de Mariapaola d'Imperio, la courbe de fréquence fondamentale demeure en revanche stylisée en cibles H et L (choix réaffirmé lors de son exposé d'Habilitation à diriger des recherches, Université Paris 3-Sorbonne Nouvelle, juin 2005).

montre à l'évidence qu'il s'agit en fait d'un ensemble de phénomènes intonatifs et non lexicaux :

> Le contexte forge le schéma accentuel d'un mot. La chaîne parlée est la base sur laquelle se forme la configuration du schéma lexical. Ainsi, on ne peut pas étudier l'accent lexical isolément ; il faut l'étudier dans son emploi réel, dans la chaîne parlée. (Hoa 1980:260)
>
> Si on excepte les particules grammaticales atones, on ne peut, comme le font certains ouvrages, statuer sur l'accentuation d'une syllabe ou d'un mot isolément ; il faut l'étudier dans un contexte, contexte au sein d'une phrase et contexte plus large (linguistique et extralinguistique). Il faut examiner la séquence globale, dans un contexte global. (Hoa 1980:264)

L'emploi de l'adjectif « lexical » par l'auteur vise sans doute à mettre en avant le fait que la fréquence d'emploi d'un mot, et sa signification, jouent un rôle dans ces processus, qui de ce fait ont partie liée avec le lexique ; néanmoins, **il s'agit sans conteste de processus intonatifs**, comme le fait par ailleurs ressortir la représentation proposée par l'auteur pour figurer la place de l'accent à l'intérieur du domaine de la prosodie. Cet emploi original du terme « accent lexical » pourrait induire en erreur. (La figure 4.2 met en regard la représentation des composantes de la prosodie par Monique Hoa et par Jacqueline Vaissière.)

Ce point de vocabulaire mis à part, l'étude par Monique Hoa du découpage en constituants, du rythme et des effets intonatifs de la fréquence d'emploi d'un mot fournit une abondance d'observations sur les tensions qui travaillent un système tonal, et la variété des unités qui entrent en jeu, confrontant les modèles proposés par Liberman et Prince (Liberman 1975, *The intonational system of English ;* Liberman et Prince 1977, « On stress and linguistic rhythm » ; ainsi que Selkirk 1980) avec des données du mandarin de Pékin. Reprenant plusieurs des idées de Monique Hoa, Chen 2000 propose de décrire la « structure accentuelle » du mandarin de la façon suivante :

1) sur la première ligne, celle des syllabes, une croix est assignée à chaque syllabe
2) sur la deuxième ligne, chaque syllabe reçoit à nouveau une croix, à l'exclusion des syllabes au ton léger. A ce niveau, chaque syllabe porteuse de ton constitue une unité
3) des pieds *(feet)* sont formés en assignant une croix par syntagme.

Les lignes supérieures constituent des groupements de syntagmes à un niveau plus élevé.

Les règles qui président à la création des groupements sont complexes, de l'aveu de l'auteur, soulevant des problèmes similaires à ceux rencontrés par le modèle de Liberman pour l'anglais (Kenstowicz 1994:555 et références citées). Plus fondamentalement, la question se pose de savoir s'il existe un principe général qui commande l'assignation de l'accent en début ou en fin de groupe. Chao Yuen-ren 1968 juge que la position forte (« accentuée »)

est la position finale de groupe (idée d'une structure *iambique* ou *anapestique*). Ce point de vue est contradictoire avec l'idée sur laquelle repose le système développé par Shih Chilin et Kochanski 2000 pour la synthèse du mandarin : ce système, plus pratique que théorique, repose sur l'idée d'une structuration *trochaïque* et non iambique, les syllabes initiales ayant un coefficient de *force (strength)* plus élevé. Sur un plan théorique, Duanmu San 2000 (*The Phonology of Standard Chinese*) s'inscrit pareillement en faux par rapport à Chao Yuen-ren, défendant l'idée selon laquelle l'énoncé en chinois mandarin est structuré en pieds accentuels à tête initiale ; ce point de vue est maintenu dans un article ultérieur (Duanmu San 2004) en réponse aux critiques exposées dans le même volume par Dell 2004 (« On recent claims about stress and tone in Beijing Mandarin »), qui soulignait certaines faiblesses de l'argumentation phonologique développée par Duanmu San[1]. Ces points de vue antagonistes suggèrent que les faits ne parlent pas d'eux-mêmes ; les remarques de Moira Yip nous amènent à penser que ni l'un ni l'autre de ces deux points de vue n'est adéquat (Yip 1993:490) :

> ... native speakers tend to reject the idea that there is any detectable stress (...), and data collected in the phonetics lab at the Academy of Social Sciences in Beijing found no consistent judgments on the relative stress of fully toned syllables (Wang Hongjun, p.c.).

Monique Hoa, plus nuancée, parle de *tendances :* la tendance générale serait *iambique* dans les séquences verbe + objet, *trochaïque* dans les séquences déterminant + déterminé. En outre, les mots les plus courants auraient tendance à suivre un schéma trochaïque plutôt qu'iambique. Enfin, le ton 3 (le plus bas des quatre) serait moins susceptible de recevoir un « accent » (remarque compatible avec l'idée de Meredith 1990 selon laquelle le ton haut et le ton descendant attireraient un « accent »). Monique Hoa note l'écart entre le mode de calcul qu'implique la grille métrique et le détail des observations, et en conclut de façon catégorique : « En conclusion, la grille métrique n'a pas d'utilité pour le chinois mandarin » (*op. cit.*, p. 279).

[1] Les mêmes arguments s'appliquent à l'ouvrage (en chinois) de Wang Hongjun 王洪君 (1999), Hanyu feixianxing yinxixue 汉语非线性音系学, 北京大学出版社. Une bibliographie des analyses du chinois dans le cadre métrique est proposée par Chen 2000.

Au bilan, il apparaît que les *degrés d'accent* que distingue Monique Hoa sont déterminés par de multiples facteurs ; s'y rejoignent les modules désignés ici comme intonation syntaxique et intonation pragmatique, ainsi que le rythme. Tout en insistant, par souci de clarté typologique, sur le fait que ces *degrés d'accent* ne sont pas lexicaux, il faut souligner que la description de Monique Hoa fait ressortir les tensions qui travaillent la prosodie du mandarin, et aide par là à se représenter les chemins par lesquels un système tonal peut évoluer vers un système de type accentuel[1].

[1] Ce n'est pas ici le lieu de discuter plus avant des questions que soulève la description de l'intonation du mandarin, non plus que d'autres dialectes chinois ; nous nous contentons de renvoyer aux Actes du Congrès international des sciences phonétiques (ICPhS) et du colloque Speech Prosody, qui contiennent des communications relativement abondantes au sujet du chinois ; au sujet du cantonais, signalons la thèse de Choi-Yeung Flynn (née Chang) (Flynn 2003, *Intonation in Cantonese*) et références citées.

Figures du volume I

Une table des figures est proposée en page 518.

Les figures des Annexes se trouvent en pages 494-509 ; une table en est proposée page 521.

Fig. 0.1.

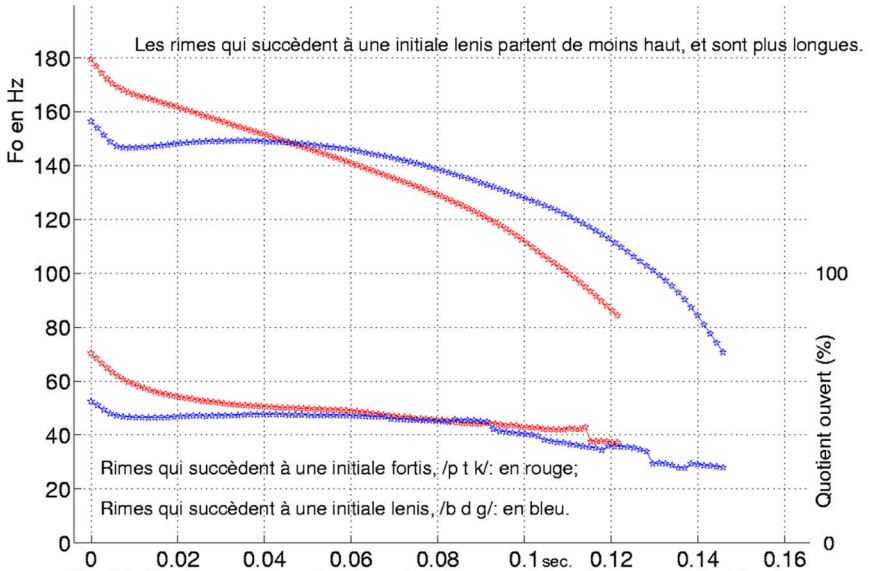


Fig. 0.1. Courbes moyennes des rimes de 192 monosyllabes; langue: anglais; 4 locuteurs

Fig. 0.2.

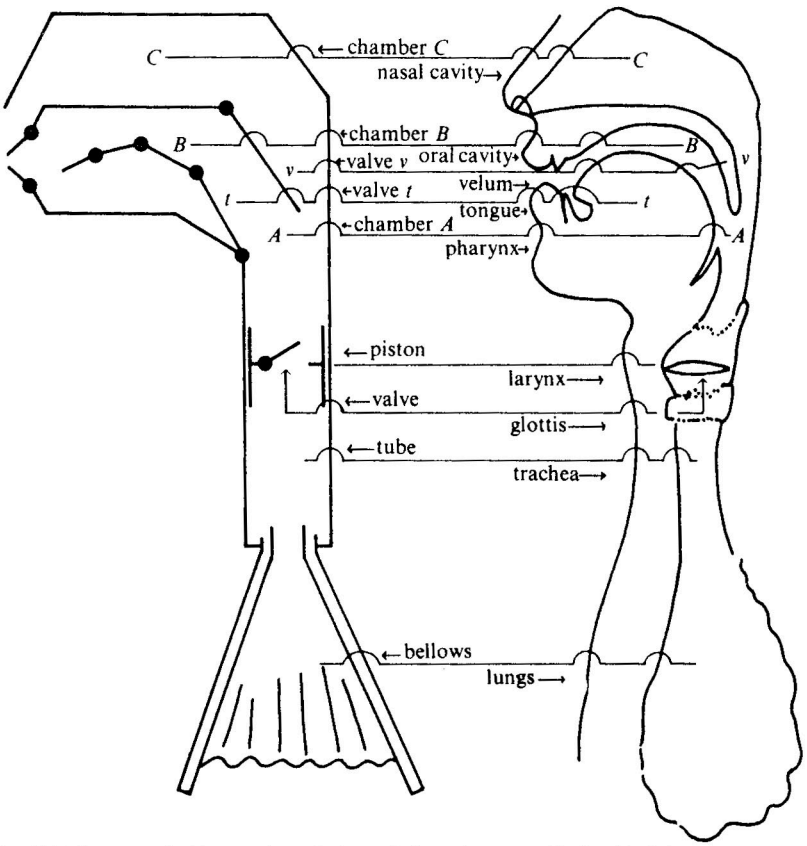


Fig. 0.2. Le conduit vocal : schéma faisant ressortir la division en trois composantes, sous-glottique, glottique et supraglottique. Repris de Catford (1977:18).

Exemples de positionnement des électrodes de l'électroglottographe.
(photos A. Michaud, informateurs de langue naxi)



photo 1 (locuteur M13)



photo 2 (locuteur M8)

Noter la façon dont les électrodes sont fixées à la bande velcro



photo 3 (locutrice F2)



photo 4 (locuteur M7)

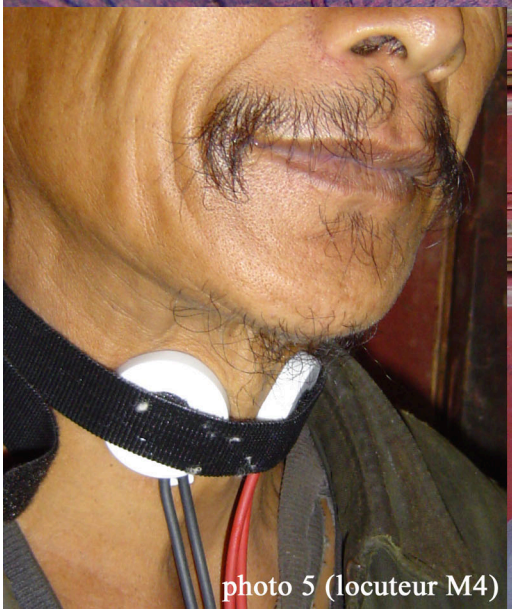
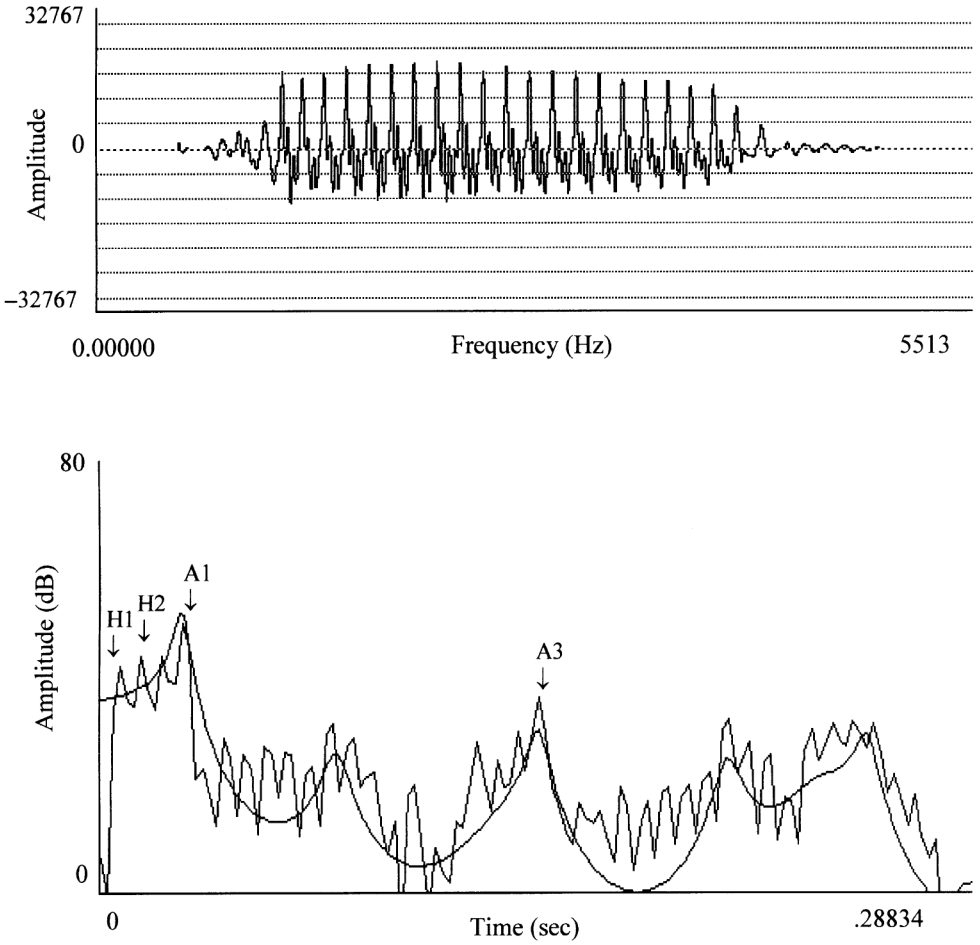


photo 5 (locuteur M4)



photo 6 (locuteur M12)

Branchements de terrain: avec DAT, carte son externe, et PC portable.



The top panel shows the waveform of the word [pɚ:p] "to run into someone" spoken by speaker W1. The bottom panel shows LPC (smooth curve) and DFT (jagged curve) spectra derived at approximately the middle of the vowel and illustrates how the amplitude of H_1 and H_2, as well as the most prominent harmonic in the F_1 and F_3 regions (A_1 and A_3), respectively, were obtained.

Fig. 1.1. Exemple de coupe spectrale, avec indication de valeurs couramment relevées pour l'estimation de la qualité de voix. Repris de Wayland et Jongman 2003:190.

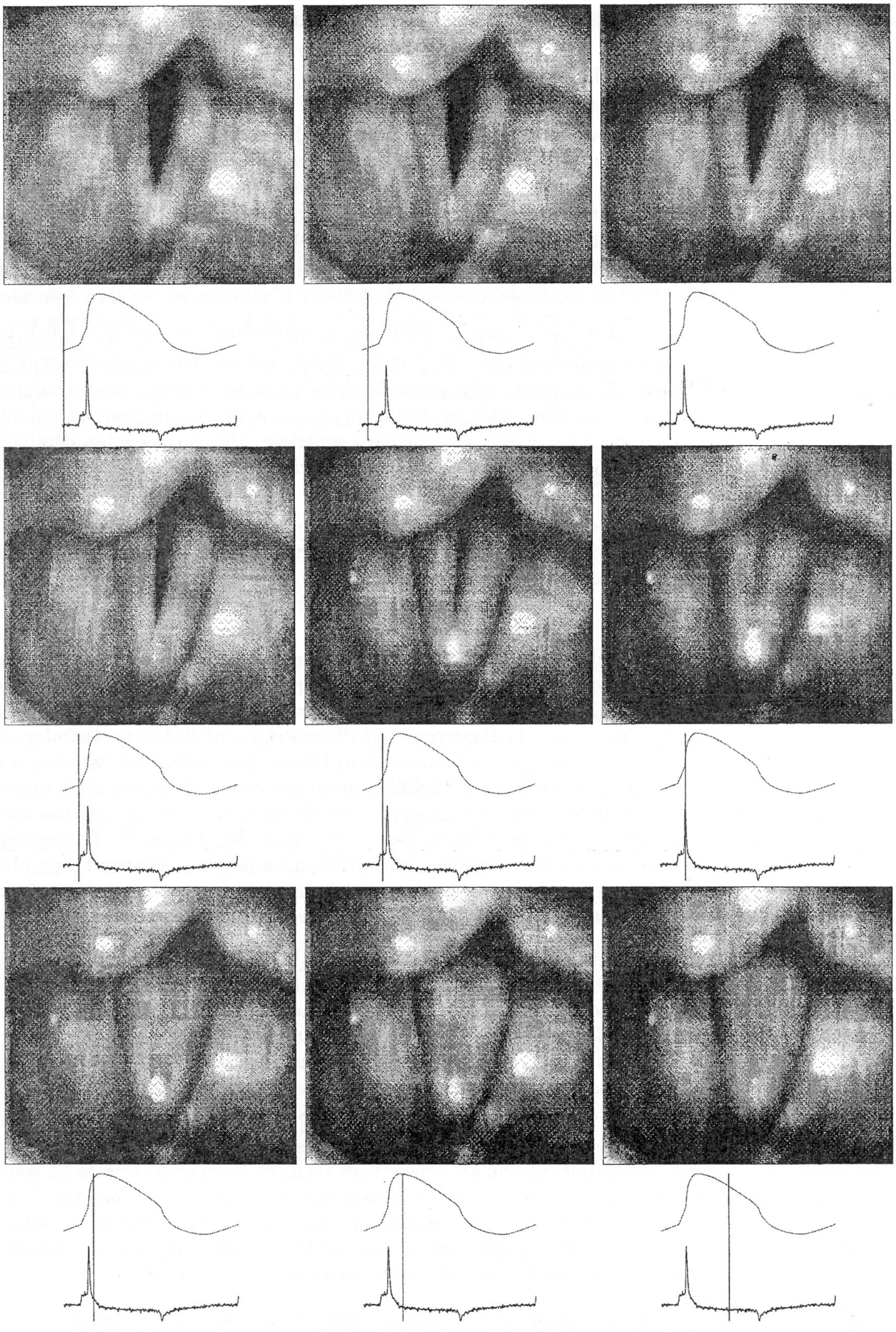


Figure 1.2 (reprise de Henrich 2001:94):

Visualisation de la fermeture *par cinématographie ultra-rapide et électroglottographie simultanées (locuteur en phonation normale,* $f_0 = 110\,Hz$*, mécanisme I).*

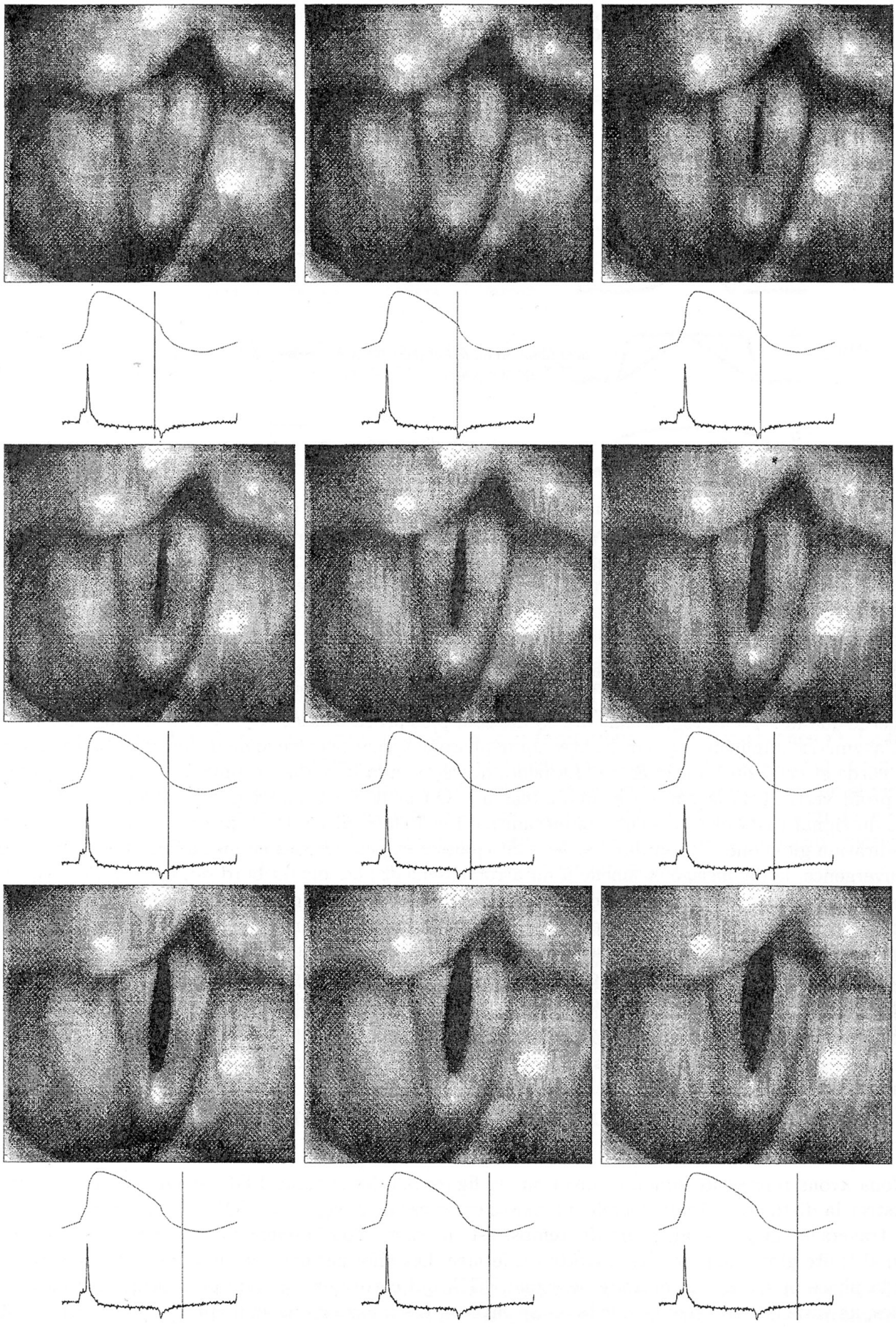
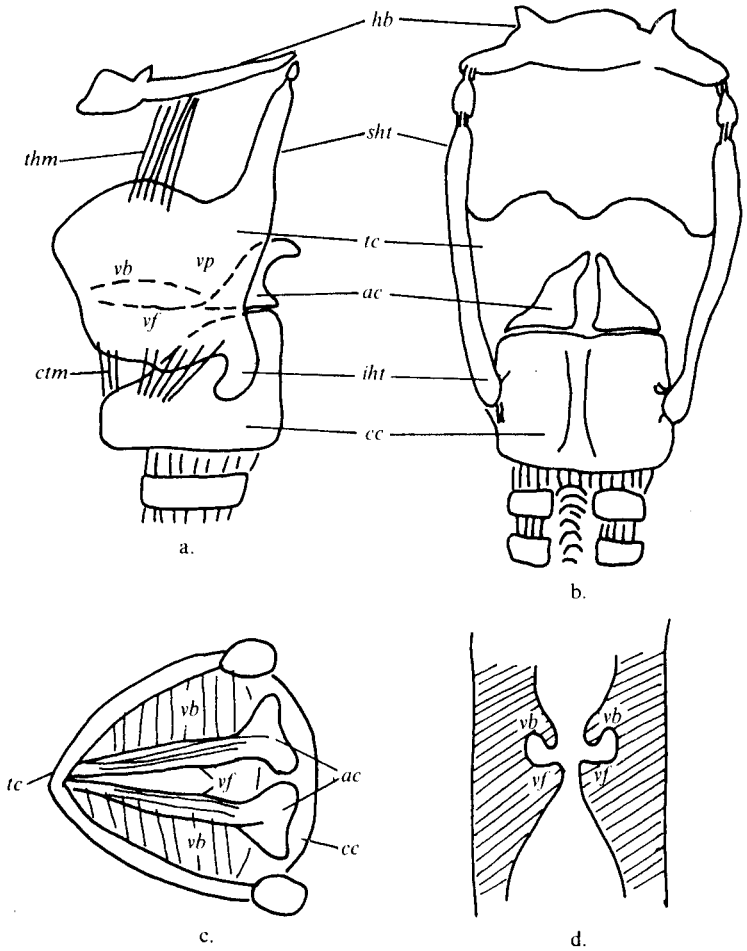


Figure 1.3 (reprise de Henrich 2001:95):

Visualisation de l'ouverture *par cinématographie ultra-rapide et électroglottographie simultanées. (locuteur en phonation normale,* $f_0 = 110\,Hz$*, mécanisme I).*



hb represents the hyoid bone in the neck from which the larynx is suspended; tc is the thyroid cartilage, which is suspended from the hyoid bone by thm, the thyro-hyoid muscle; and sht the superior horns of the thyroid cartilage. The thyroid cartilage articulates by means of its inferior horns, iht, with the sides of the cricoid cartilage, cc. Poised on top of the swollen back part of the croicoid cartilage are the twin arytenoid cartilages, ac. Running forward from the vocal processes, vp, of the arytenoid cartilages to the inside of the front wall of the thyroid cartilage are the vocal folds, vf. Just above and parallel with the vocal folds is an additional pair of folds running from back to front, the 'false vocal cords' or ventricular bands, vb. The thyroid cartilage is attached to the cricoid cartilage at the front by the crico-thyroid muscle, ctm, which can act to tilt the thyroid cartilage forward and thus stretch the vocal folds.

Figure 1.4. Le larynx vu (a) du côté gauche, (b) de l'arrière, (c) de dessus, (d) en section transversale. Repris de Catford (1977:94).

Fig. 1.5.

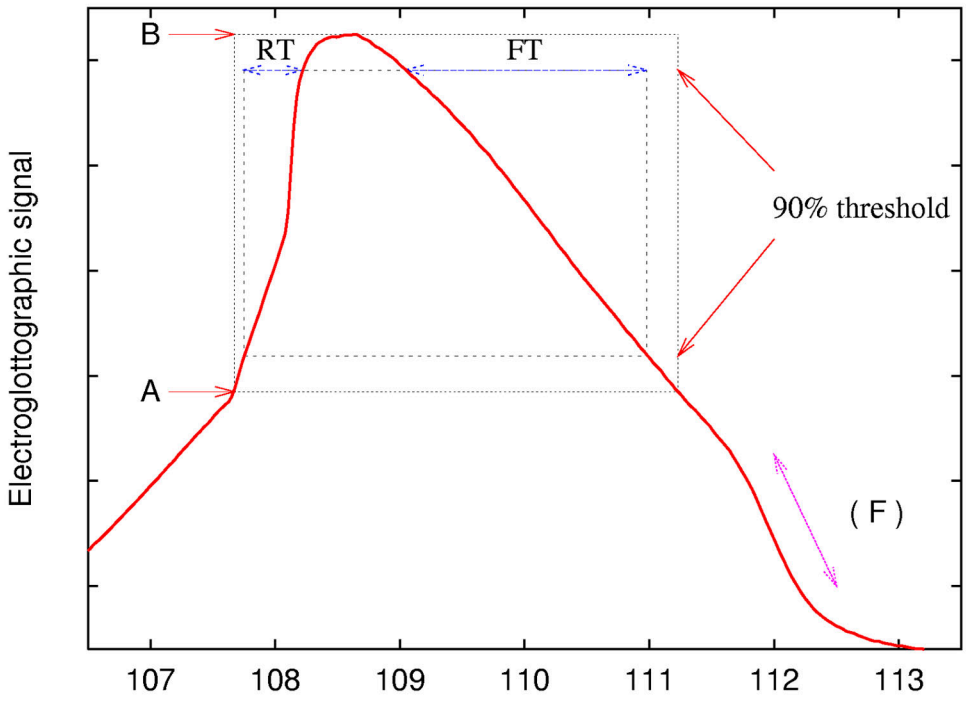


Fig. 1.5. Illustration of the method of Esling (1984). A='rapid rise begins'. B=highest point. RT='Rise Time', FT='Fall Time'. (F) = phase of fastest opening. (Time in ms.)

Fig. 1.6.

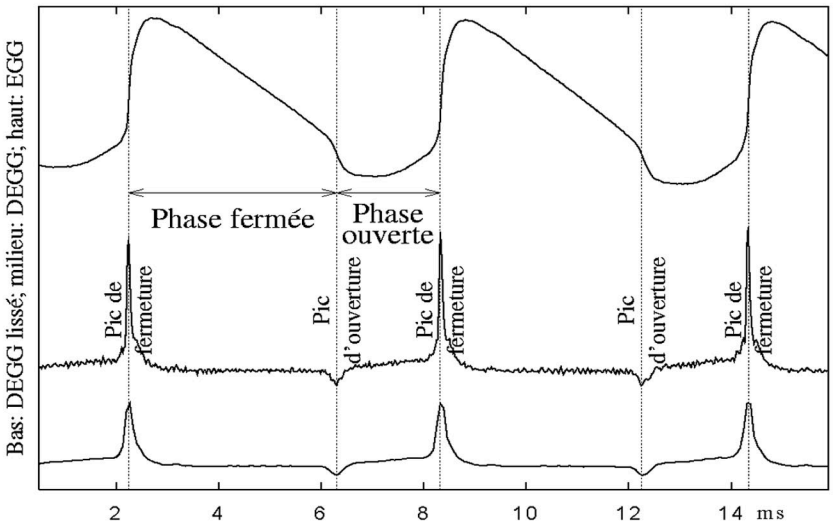


Fig. 1.6. Exemples de signaux EGG et DEGG utilisés pour le calcul de F0 et Oq: haut: EGG, milieu: DEGG, bas: DEGG lissé.

Fig. 1.7.

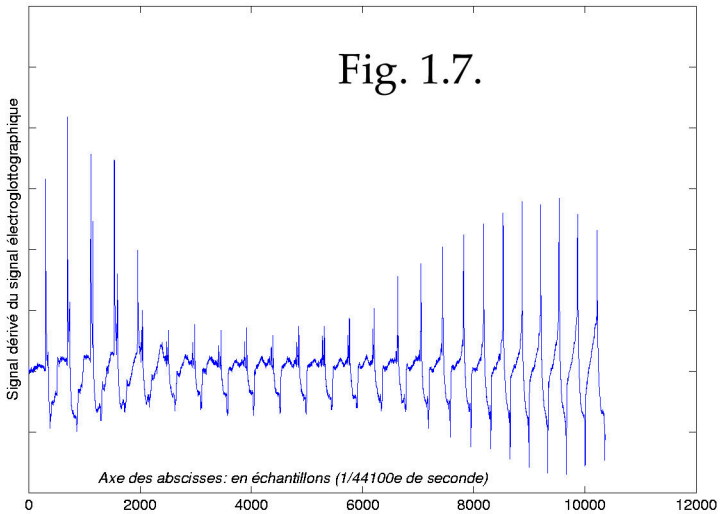


Fig. 1.7. Exemple dans lequel les pics de fermeture sont de très faible amplitude. Locuteur naxi M4.

Fig. 1.8.

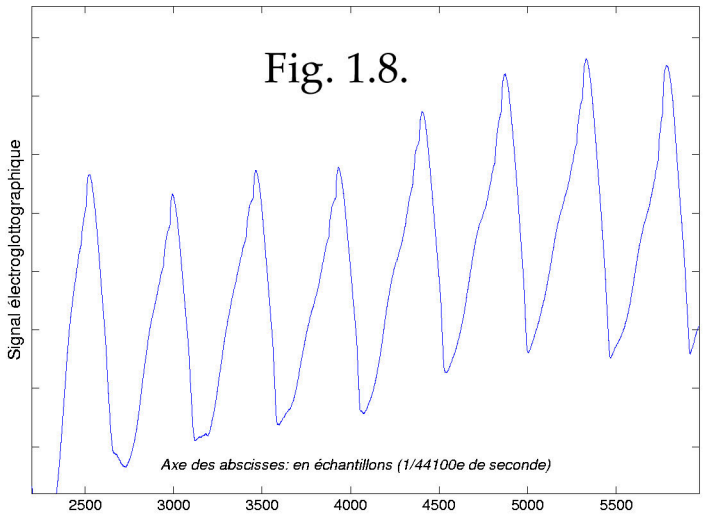


Fig. 1.8. Signal électroglottographique correspondant aux fermetures de faible amplitude de la figure 1.7.

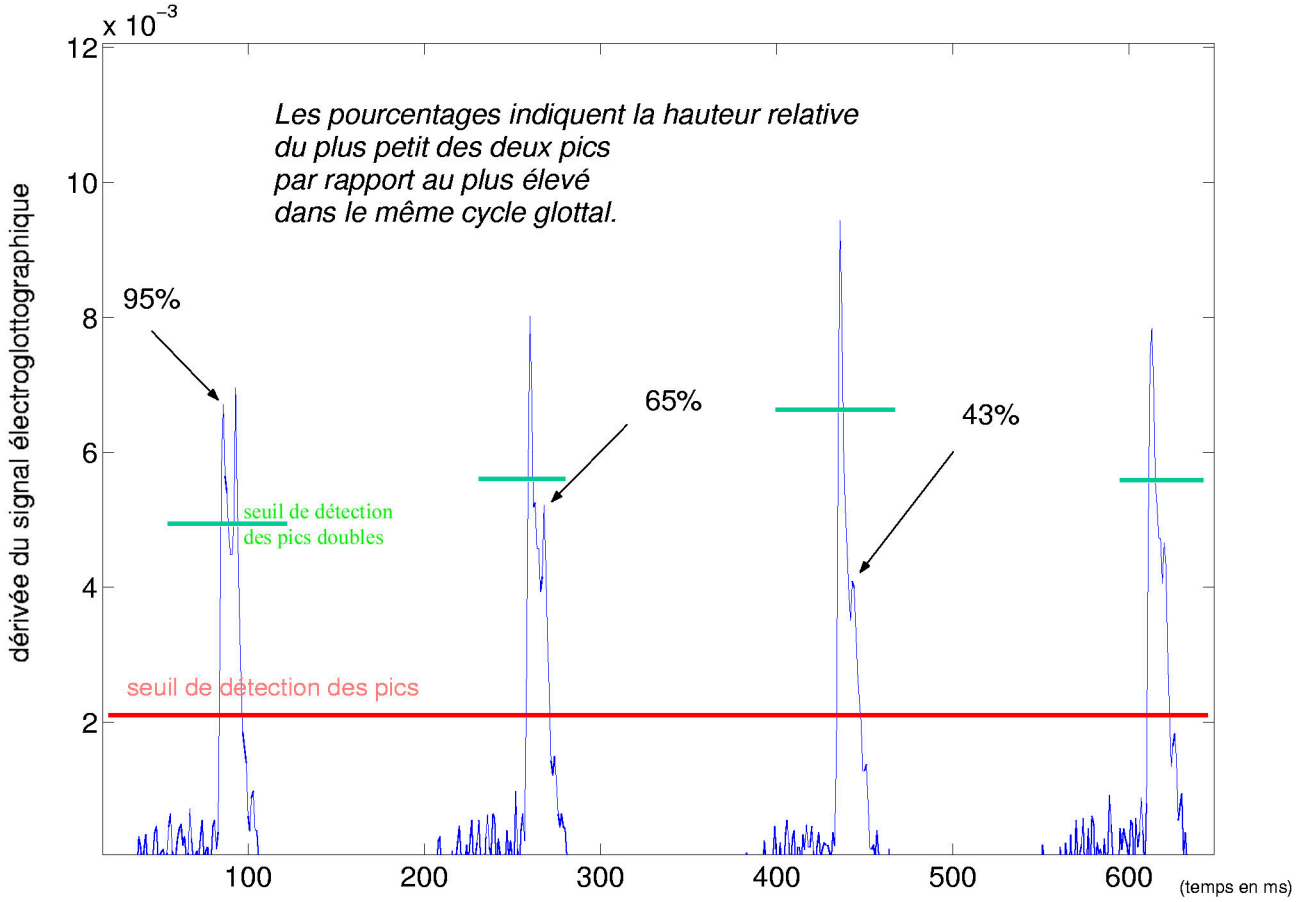


Fig 1.9. Exemple de doubles pics de fermeture. Langue: vietnamien. Locutrice F1. Début de syllabe /i/, ton A1.

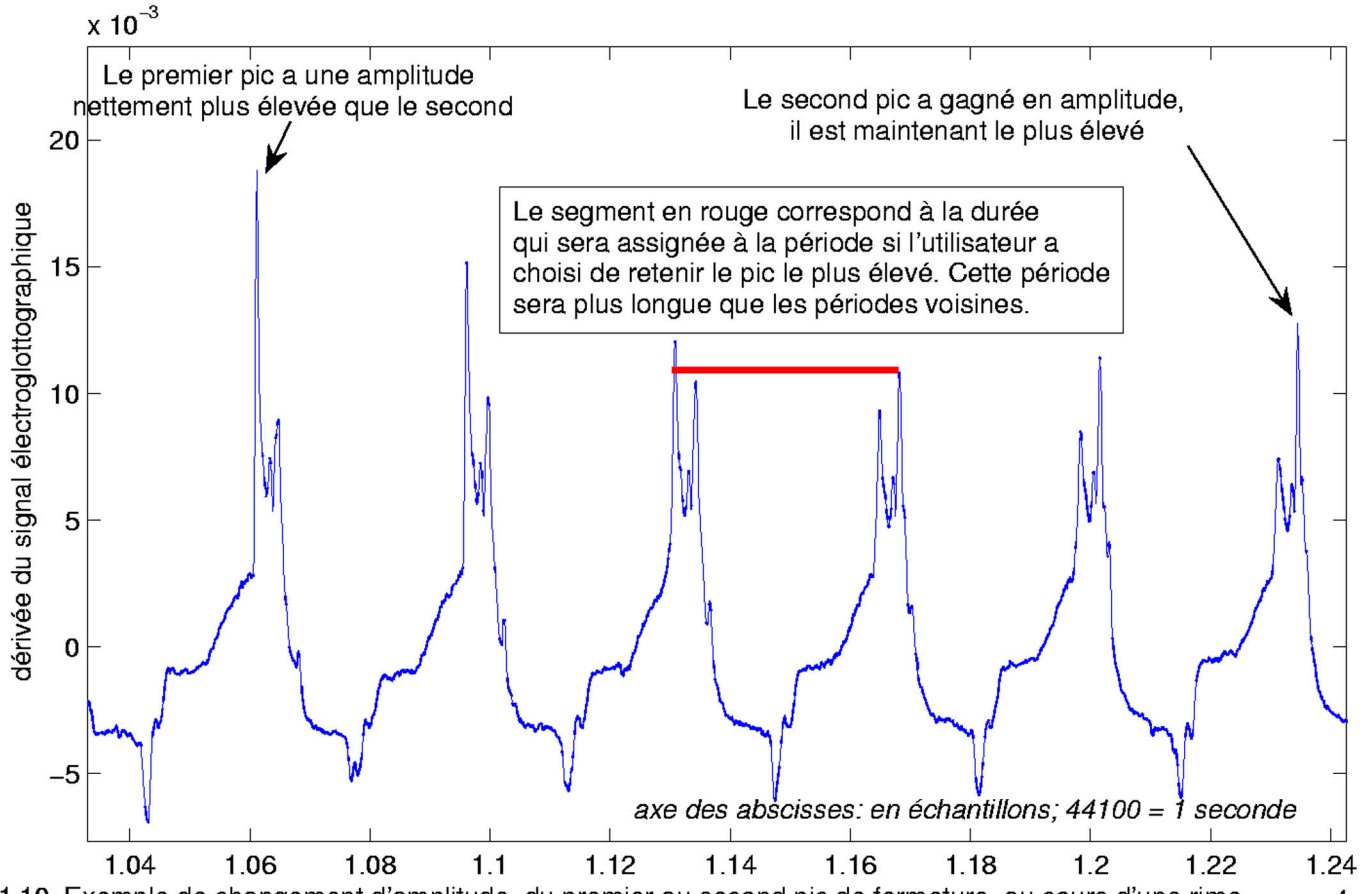


Fig. 1.10. Exemple de changement d'amplitude, du premier au second pic de fermeture, au cours d'une rime

Données de langue naxi.

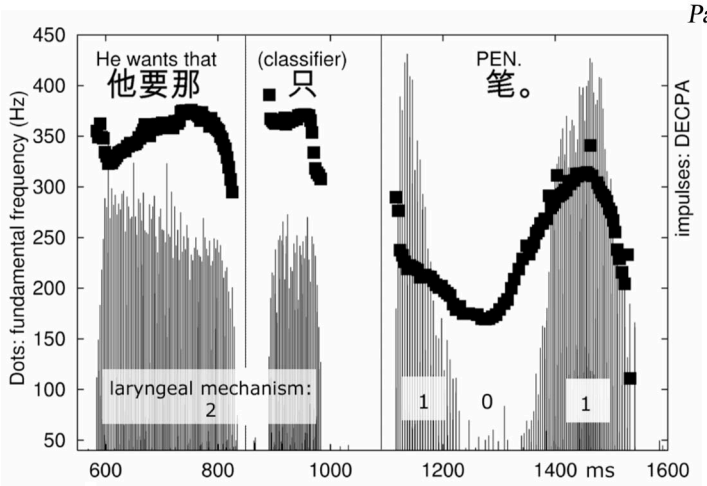


Figure 1.11. Accent on syllable 'pen' (tone 3). (Mandarin Chinese)

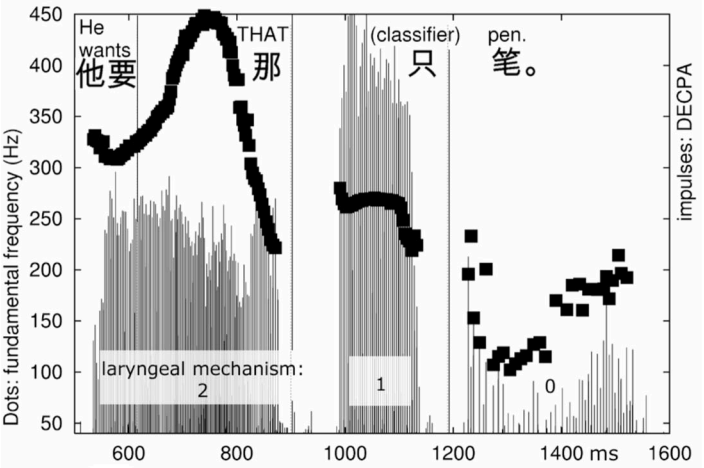
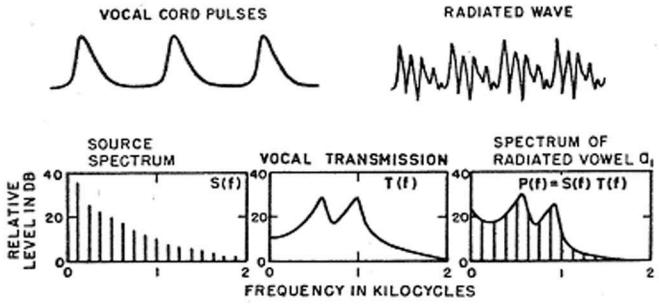


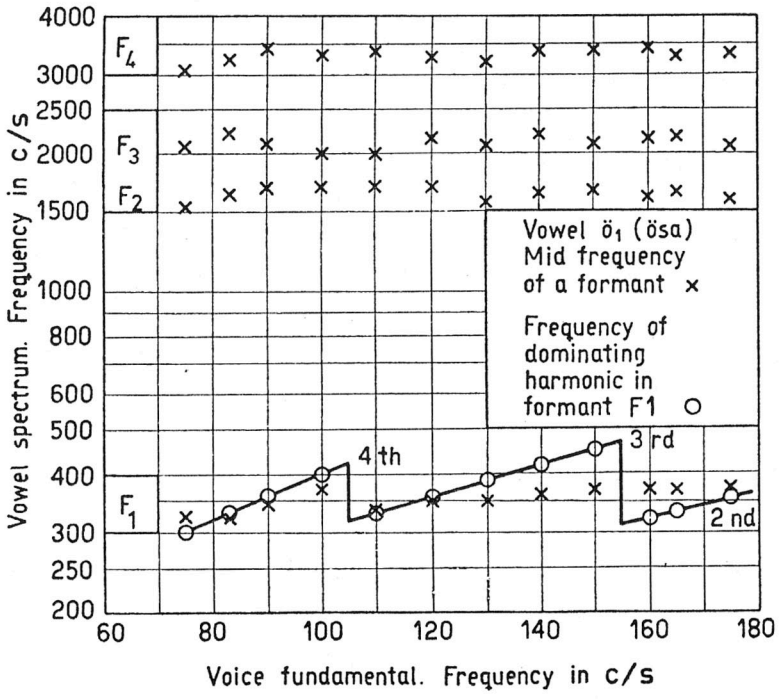
Figure 1.12. Accent on deictic (Falling tone: tone 4). (Mandarin Chinese)



Simplified source-filter decomposition of the spectrum of a two-formant voiced sound. The waveform of the periodic airflow through the glottis is converted into a harmonic spectrum *S(f)* which multiplied by the filter characteristics *T(f)* of vocal transmission provides the spectrum *P(f)* of the radiated vowel which in turn may be specified by its waveform, as seen in the upper left of the figure.

Fig. 1.13. Le modèle source-filtre de la production de la parole.
Figure reprise de Fant 1960:19.

Ci-contre : les figures 1.11 et 1.12 présentent les courbes de F0 et DECPA sur des phrases en chinois mandarin, illustrant l'effet de changements de mécanisme laryngien sur les valeurs de DECPA.



Formant frequency determinations of a vowel sustained at a series of different voice fundamental frequencies.

Fig. 1.14. Détection des formants pour une voyelle dont le timbre est maintenu, cependant que la fréquence fondamentale monte, de 75 à 175 Hz.
Figure reprise de Fant 1974:38.
Cette figure illustre la légère incertitude dans l'estimation de la fréquence de F1, qui tend à être identifiée à un harmonique, et donc à suivre F0.

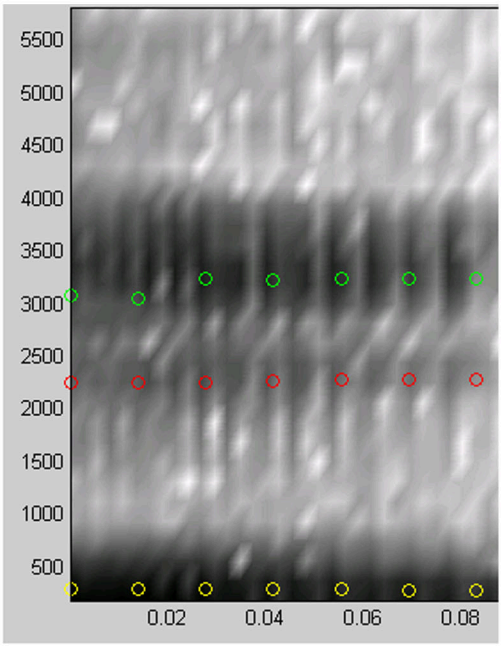


Fig. 1.15. Exemple de l'affichage des résultats de formants dans l'interface de vérification du script <forver>. La rime représentée est une voyelle anglaise /i:/.

Abscisses: temps en secondes, ordonnées: fréquences en Hz.

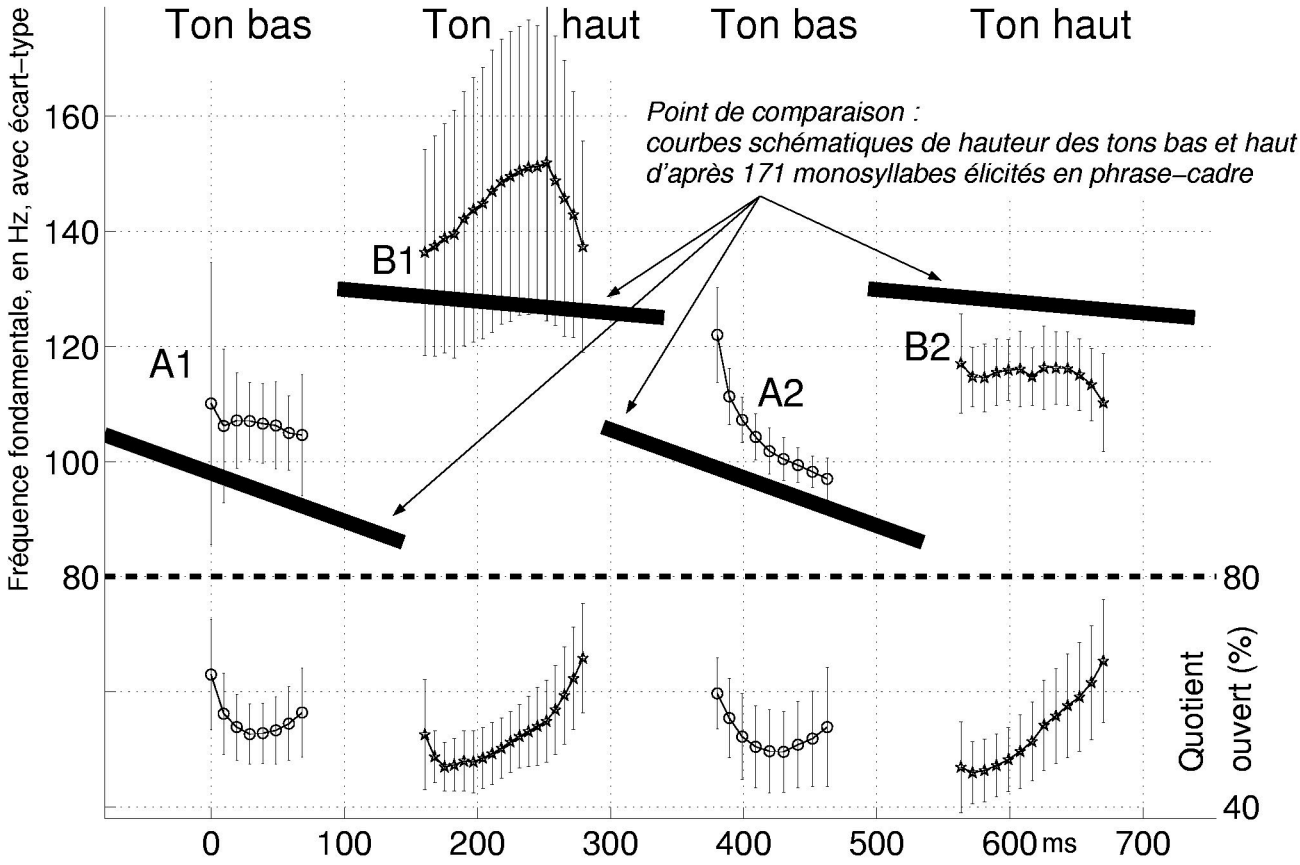


Figure 2.1a. Moyennes et écarts–type (normalisés par la durée moyenne) des tons Bas et Haut dans 10 quadrisyllabes réduplicatifs comparés aux moyennes des tons Bas et Haut sur monosyllabes en phrase–cadre (stylisés d'après 171 items, même locuteur).

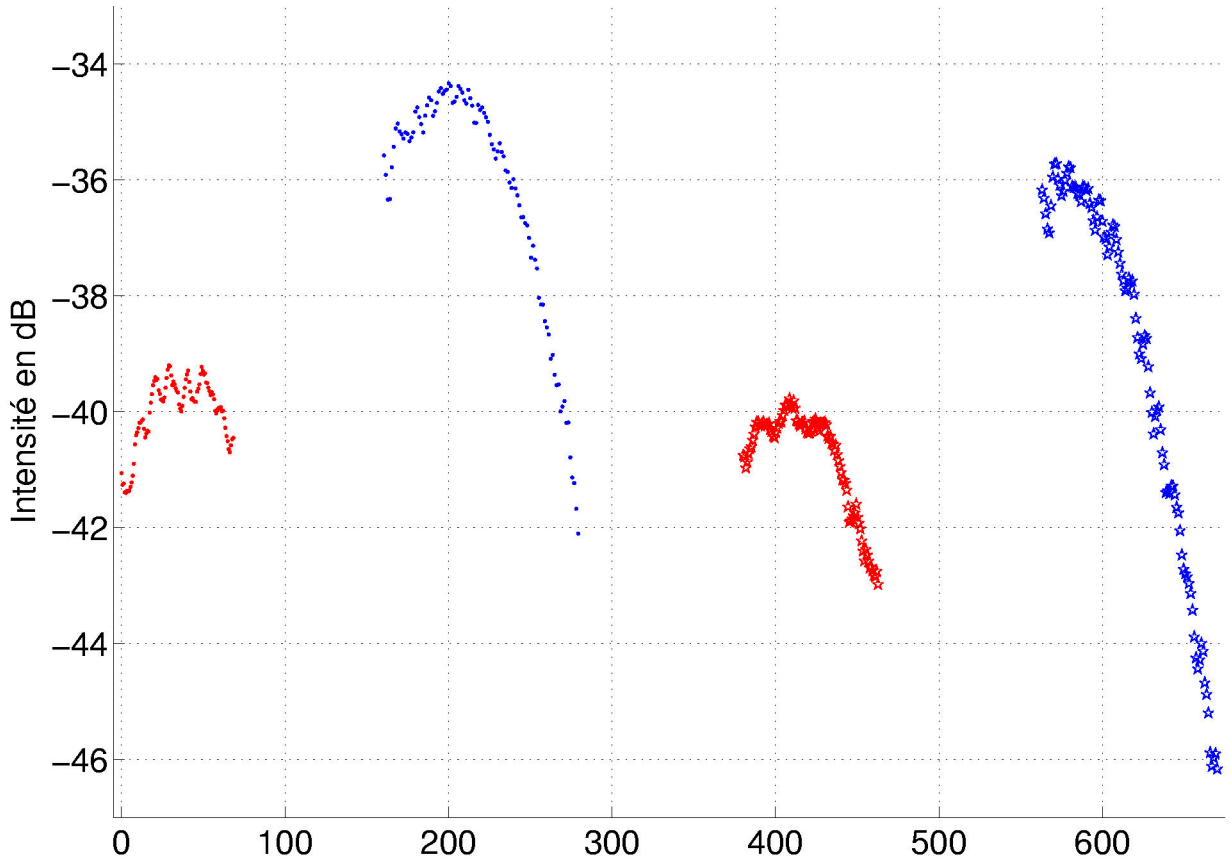
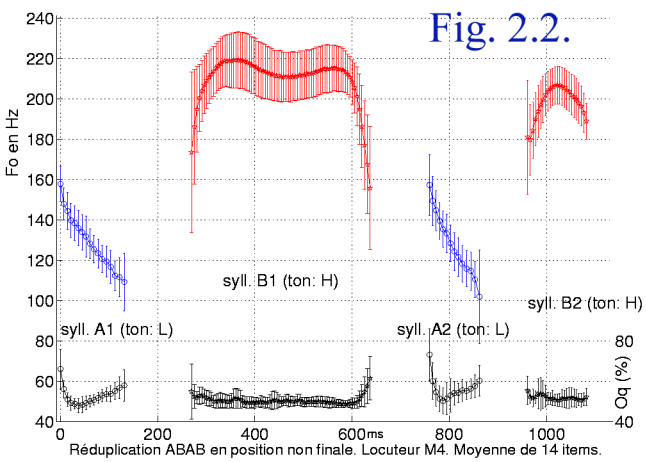
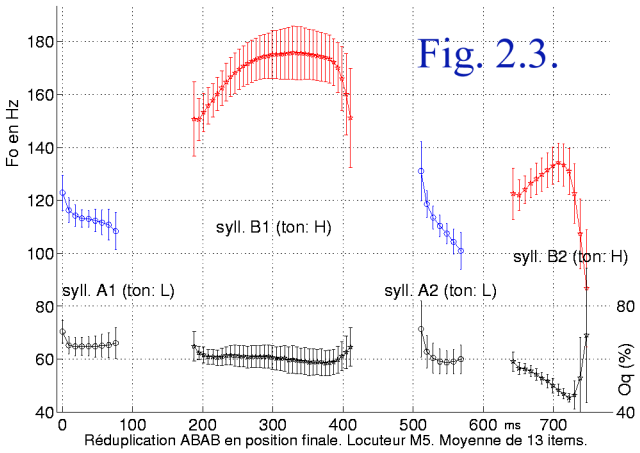
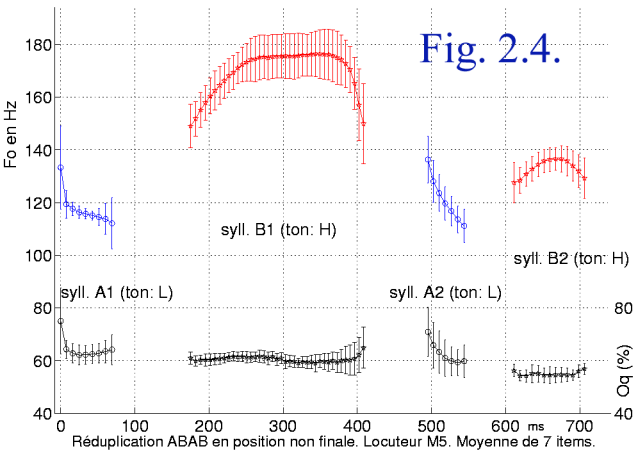


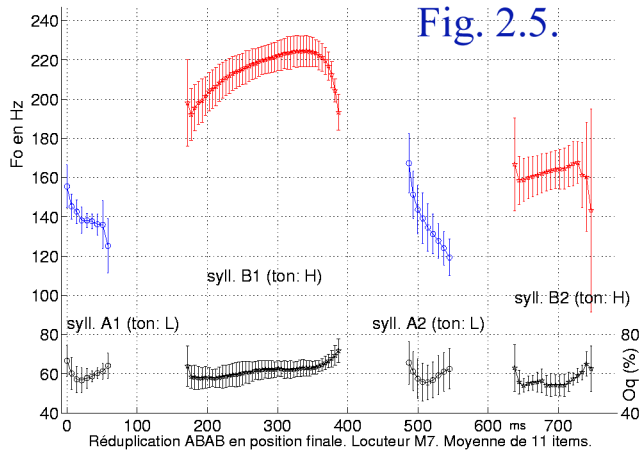
Fig. 2.1b. Courbes d'intensité correspondant aux quadrisyllabes réduplicatifs de la figure 2.1a. Réduplication ABAB en position finale. Locuteur M1. Moyenne de 10 items.

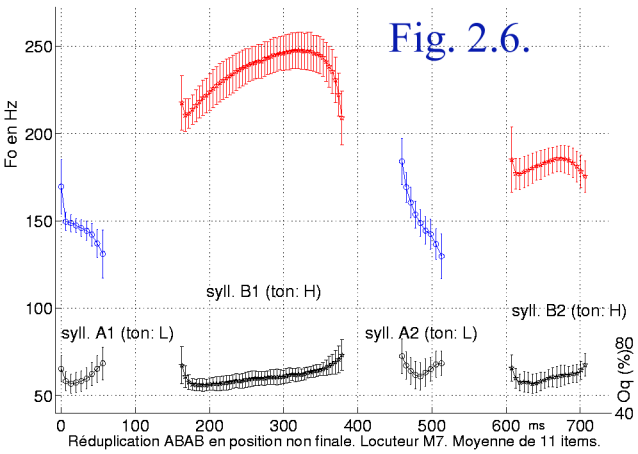
Fig. 2.2 à 2.12.
Courbes moyennes de F0 et quotient ouvert des quadrisyllabes réduplicués de schème tonal LHLH, locuteurs M4, M5, M7, M8, M9, F2. Position finale d'énoncé : nombres impairs, position non finale : nombres pairs.

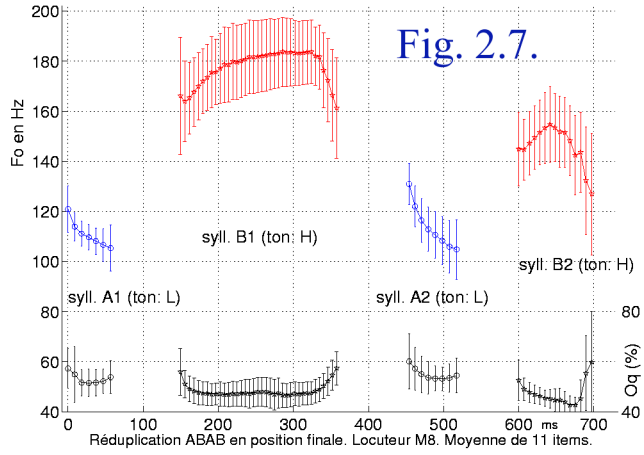












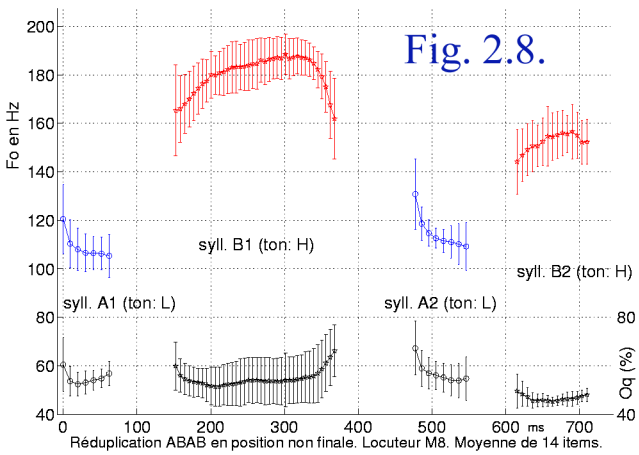
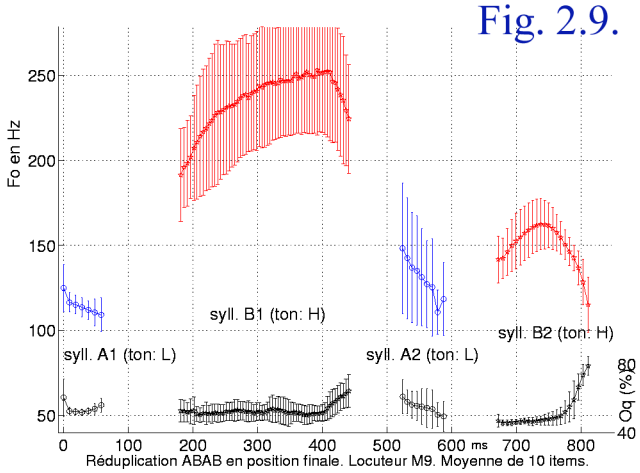
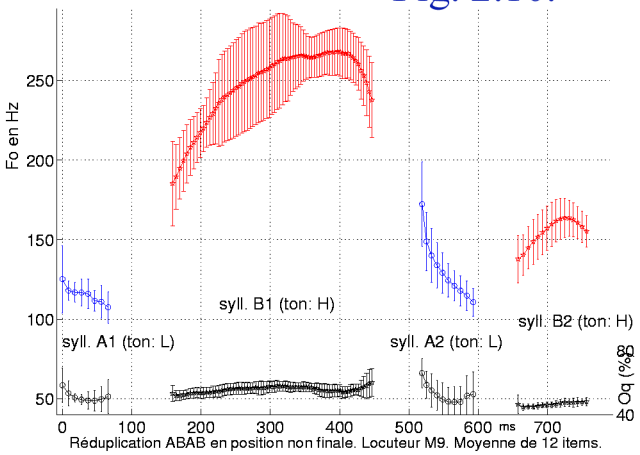


Fig. 2.9.



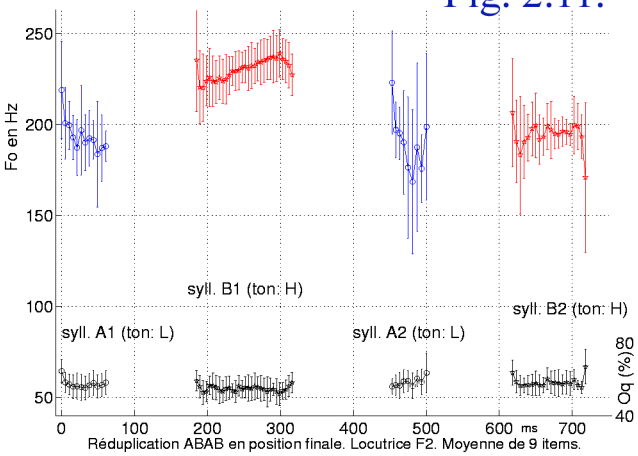
Réduplication ABAB en position finale. Locuteur M9. Moyenne de 10 items.

Fig. 2.10.



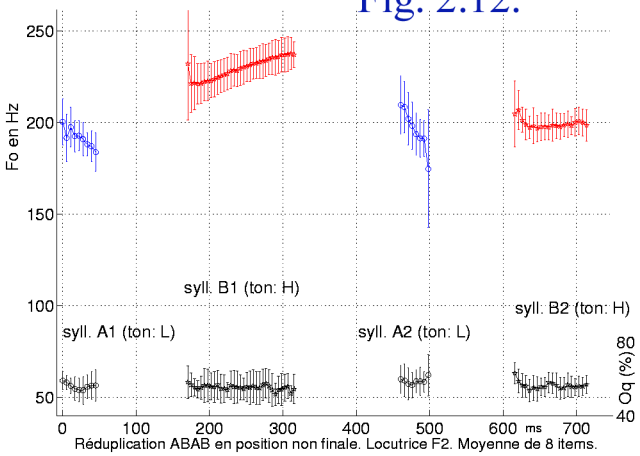
Réduplication ABAB en position non finale. Locuteur M9. Moyenne de 12 items.

Fig. 2.11.



Réduplication ABAB en position finale. Locutrice F2. Moyenne de 9 items.

Fig. 2.12.



Réduplication ABAB en position non finale. Locutrice F2. Moyenne de 8 items.

Fig. 2.13 à 2.28: courbes moyennes de F0 et quotient ouvert d'énoncés dont toutes les syllabes portent le même ton (H, M, L ou LH). Nombres impairs : courbes moyennées, nombres pairs : courbes stylisées en deux points.

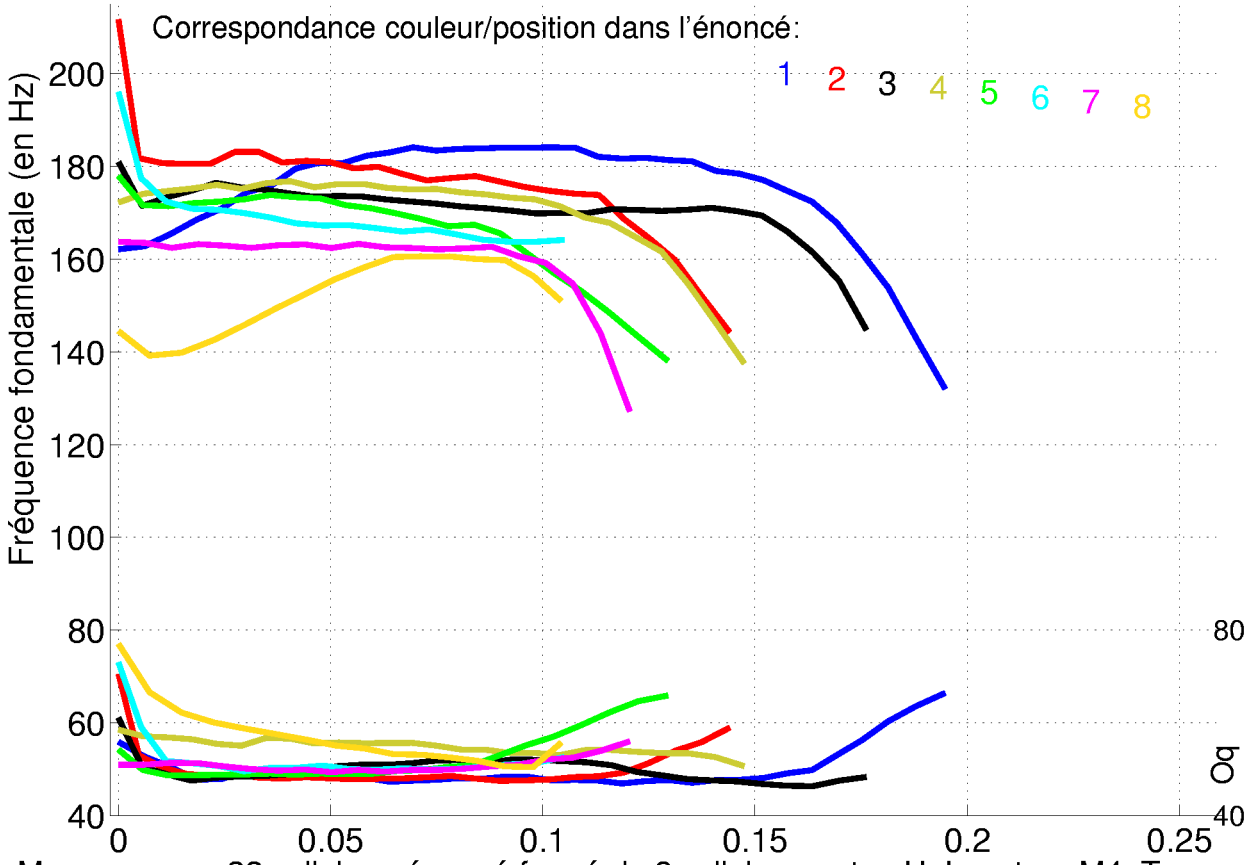


Fig. 2.13. Moyennes sur 32 syllabes; énoncé formé de 8 syllabes au ton H. Locuteur M4. Temps en s.

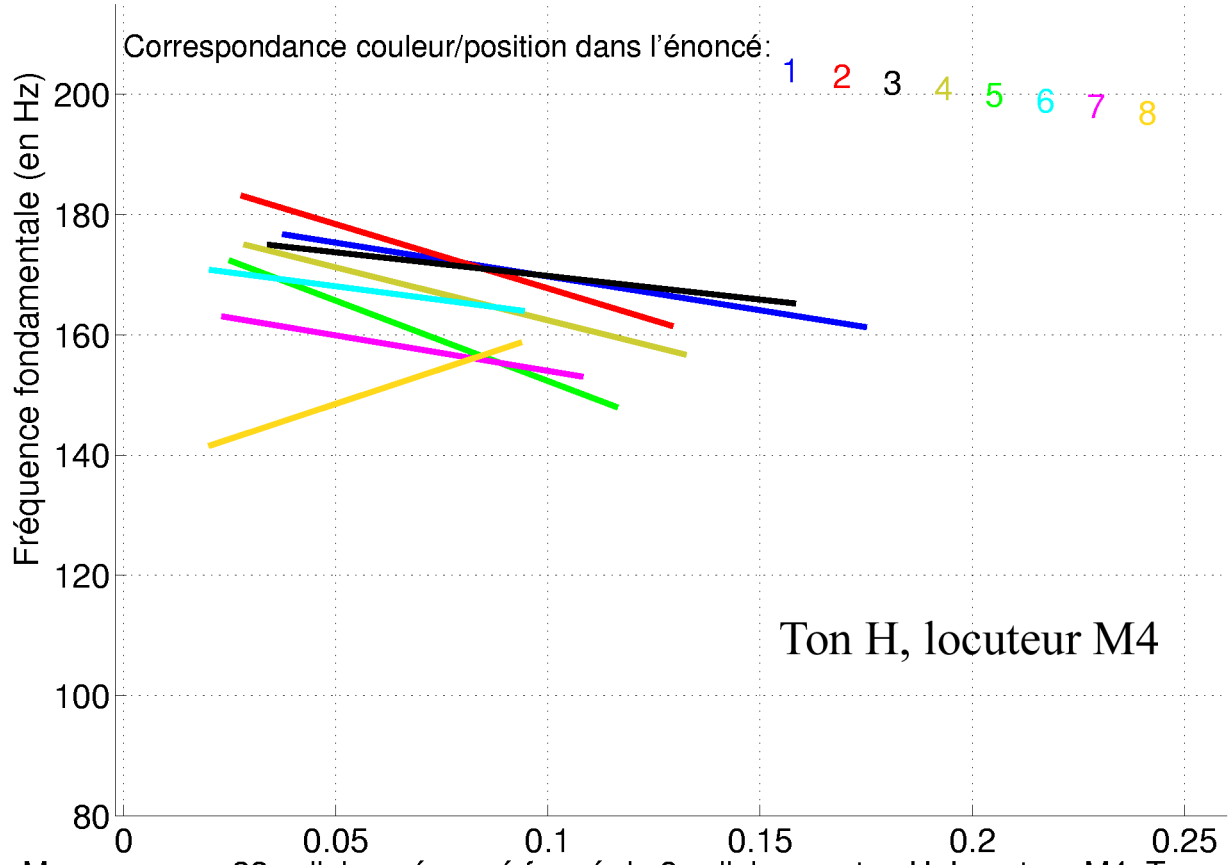


Fig. 2.14. Moyennes sur 32 syllabes; énoncé formé de 8 syllabes au ton H. Locuteur M4. Temps en s.

Fig. 2.13 à 2.28: courbes moyennes de F0 et quotient ouvert d'énoncés dont toutes les syllabes portent le même ton (H, M, L ou LH). Nombres impairs : courbes moyennées, nombres pairs : courbes stylisées en deux points.

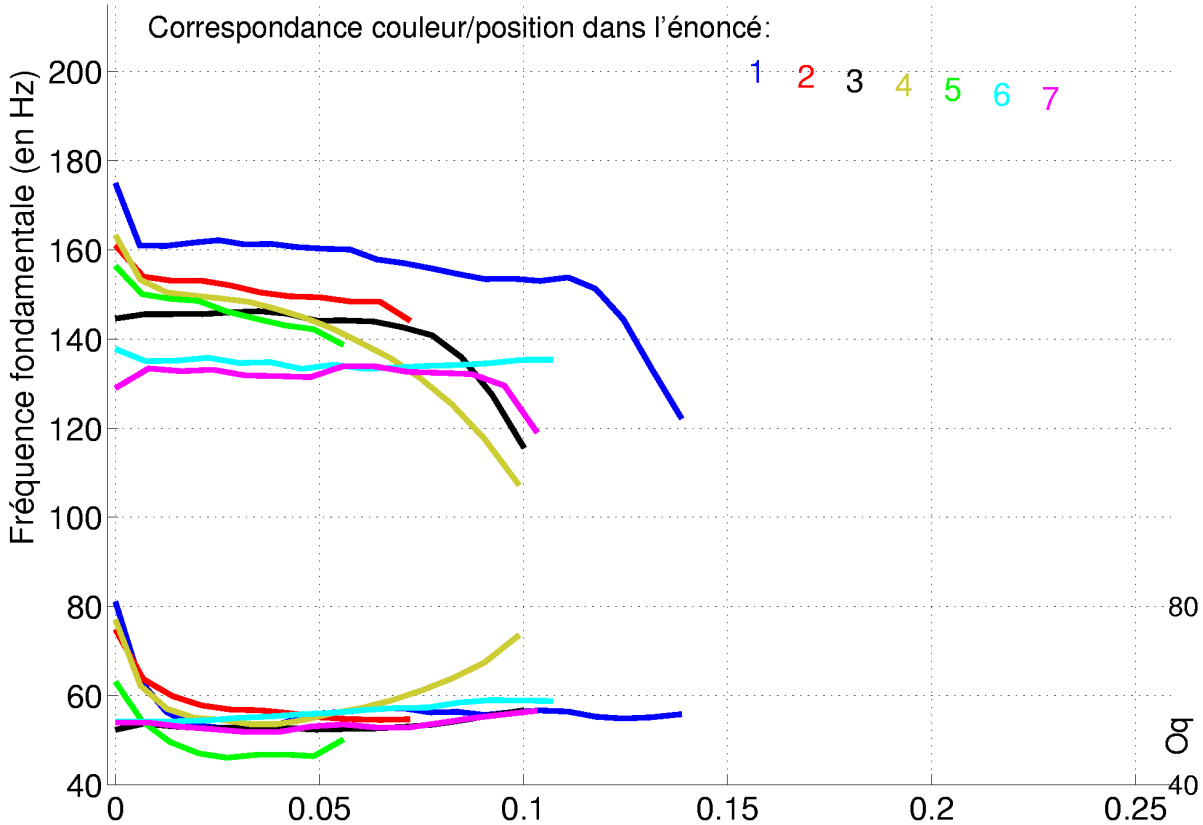


Fig. 2.15. Moyennes sur 28 syllabes; énoncé formé de 7 syllabes au ton M. Locuteur M4. Temps en s.

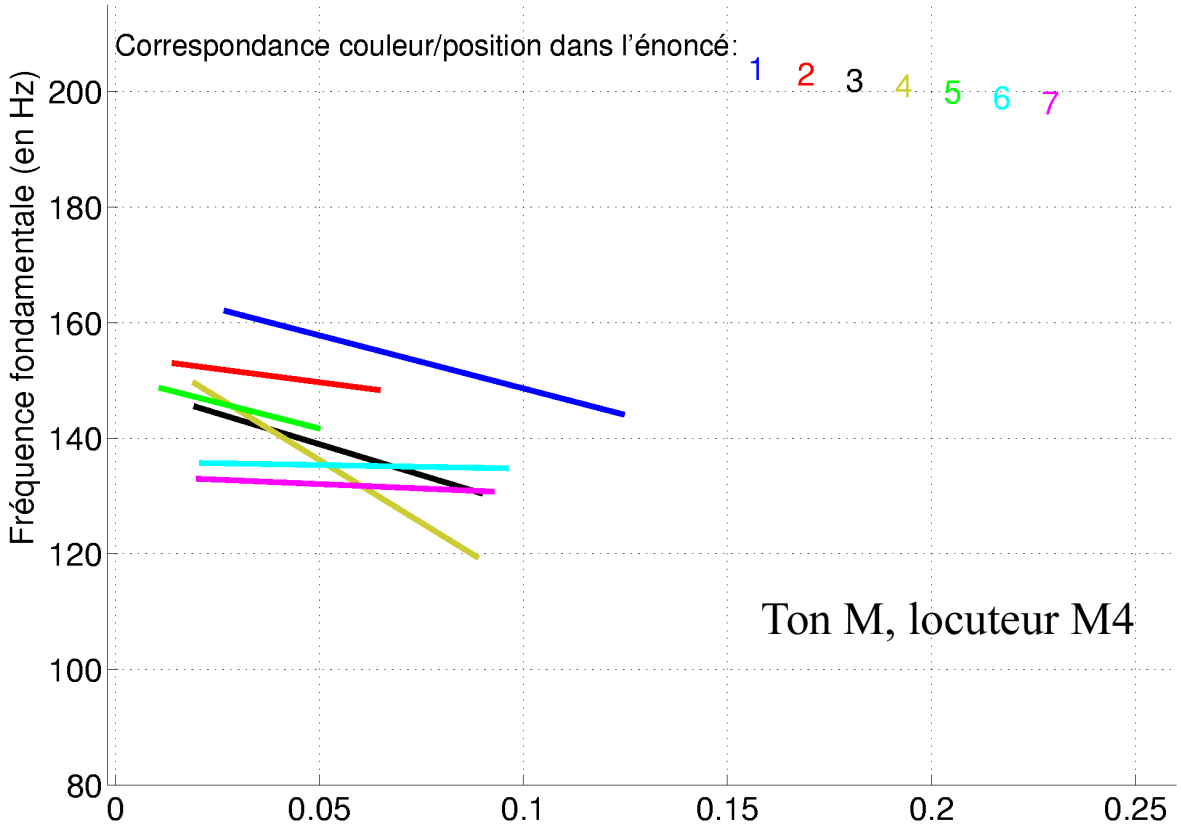


Fig. 2.16. Moyennes sur 28 syllabes; énoncé formé de 7 syllabes au ton M. Locuteur M4. Temps en s.

Fig. 2.13 à 2.28: courbes moyennes de F0 et quotient ouvert d'énoncés dont toutes les syllabes portent le même ton (H, M, L ou LH). Nombres impairs : courbes moyennées, nombres pairs : courbes stylisées en deux points.

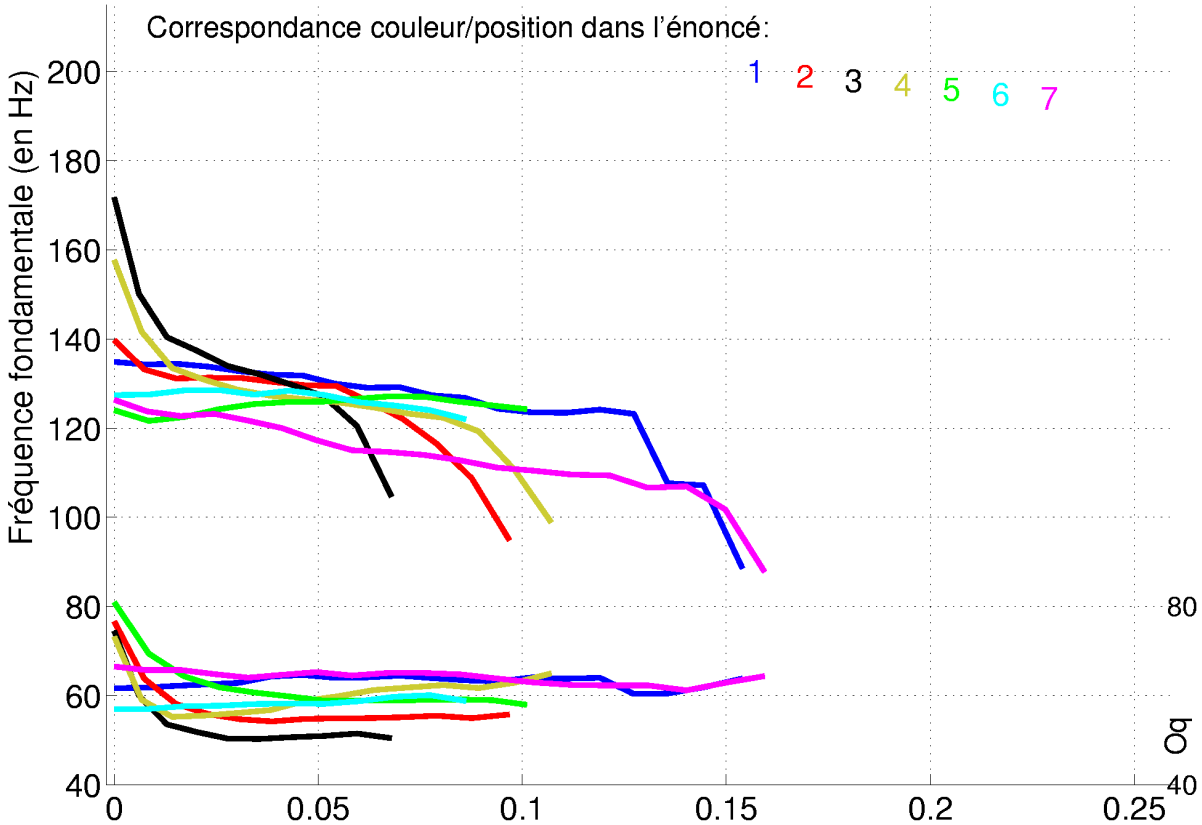


Fig. 2.17. Moyennes sur 35 syllabes; énoncé formé de 7 syllabes au ton L. Locuteur M4. Temps en s.

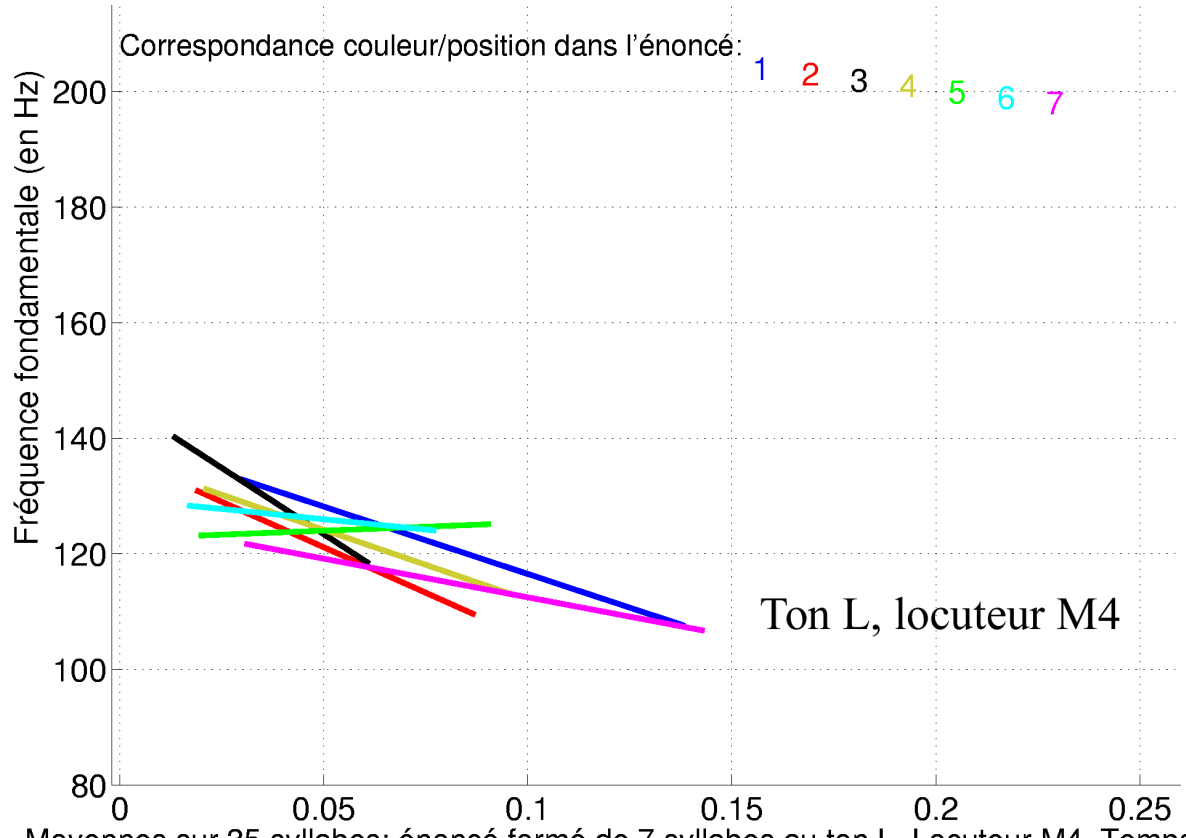


Fig. 2.18. Moyennes sur 35 syllabes; énoncé formé de 7 syllabes au ton L. Locuteur M4. Temps en s

Fig. 2.13 à 2.28 : courbes moyennes de F0 et quotient ouvert d'énoncés dont toutes les syllabes portent le même ton (H, M, L ou LH). Nombres impairs : courbes moyennées, nombres pairs : courbes stylisées en deux points.

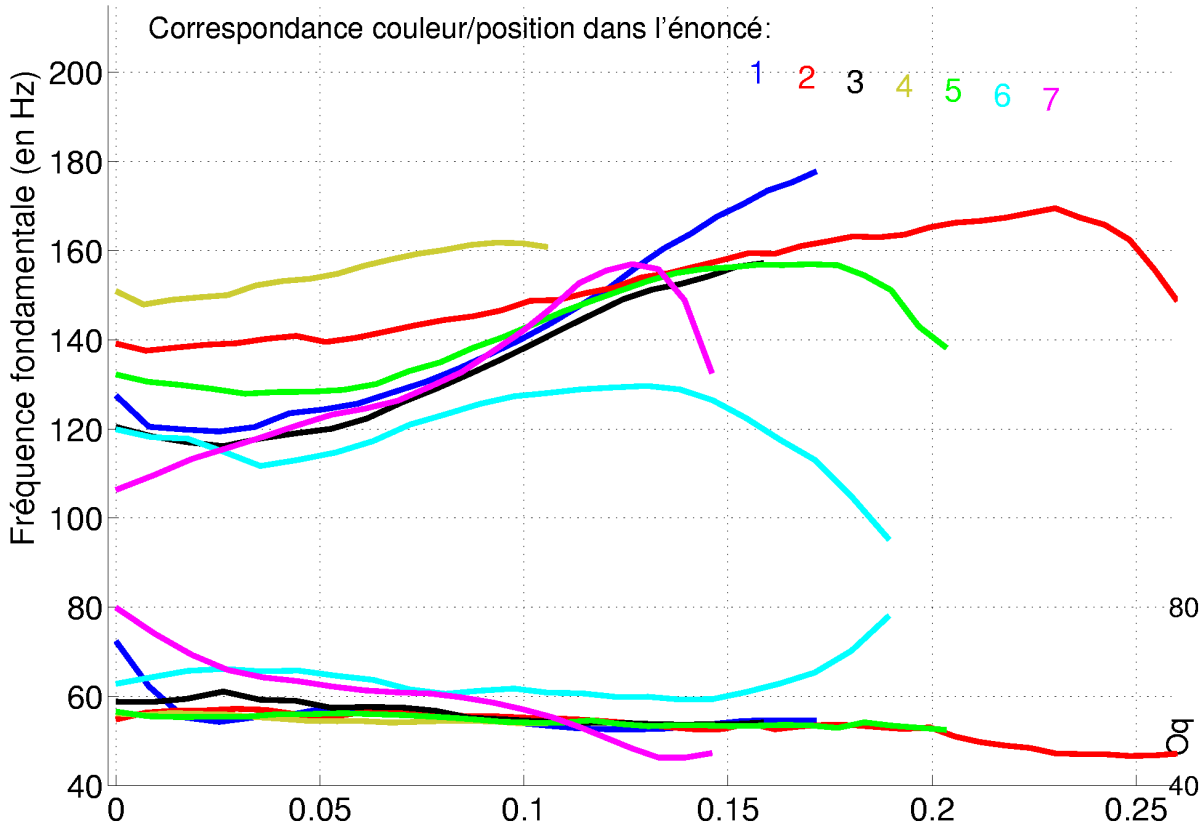


Fig. 2.19. Moyennes sur 28 syllabes; énoncé formé de 7 syllabes au ton R. Locuteur M4. Temps en s.

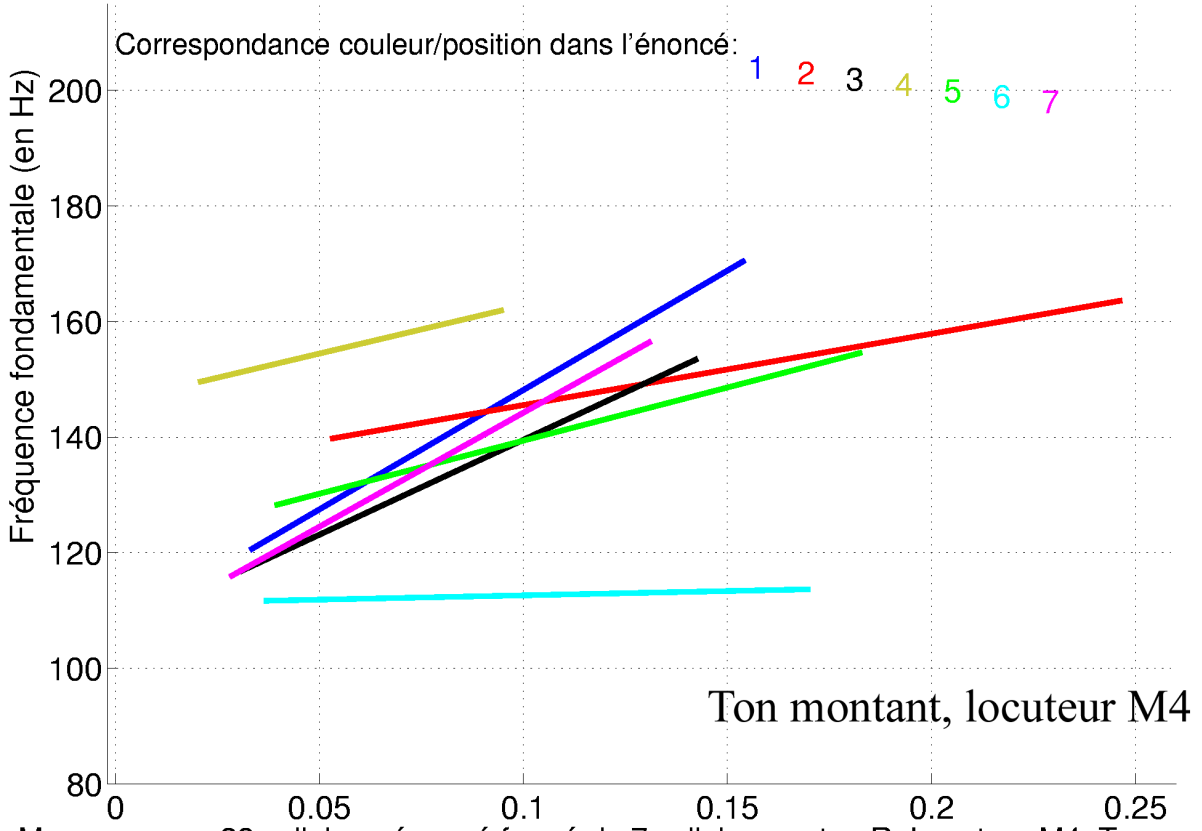


Fig. 2.20. Moyennes sur 28 syllabes; énoncé formé de 7 syllabes au ton R. Locuteur M4. Temps en s.

Fig. 2.13 à 2.28 : courbes moyennes de F0 et quotient ouvert d'énoncés dont toutes les syllabes portent le même ton (H, M, L ou LH). Nombres impairs : courbes moyennées, nombres pairs : courbes stylisées en deux points.

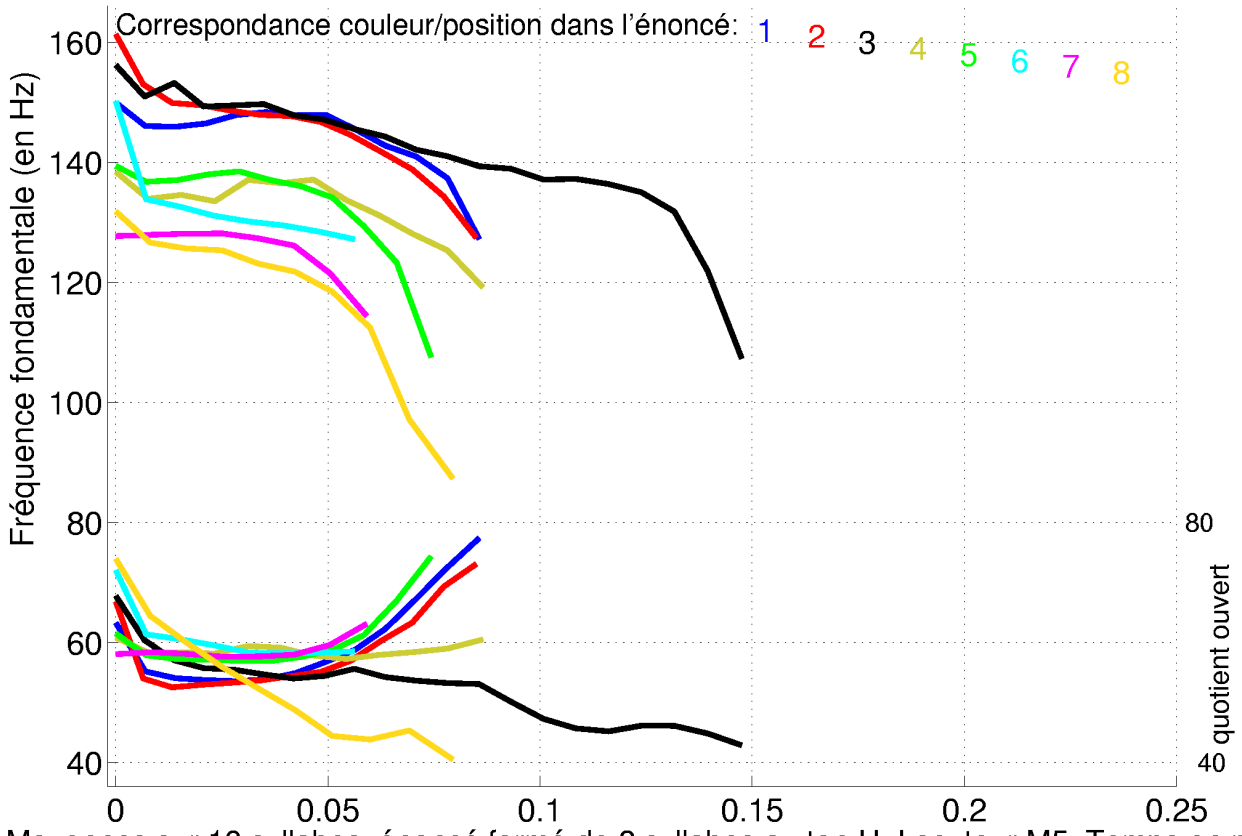


Fig. 2.21. Moyennes sur 16 syllabes; énoncé formé de 8 syllabes au ton H. Locuteur M5. Temps en ms.

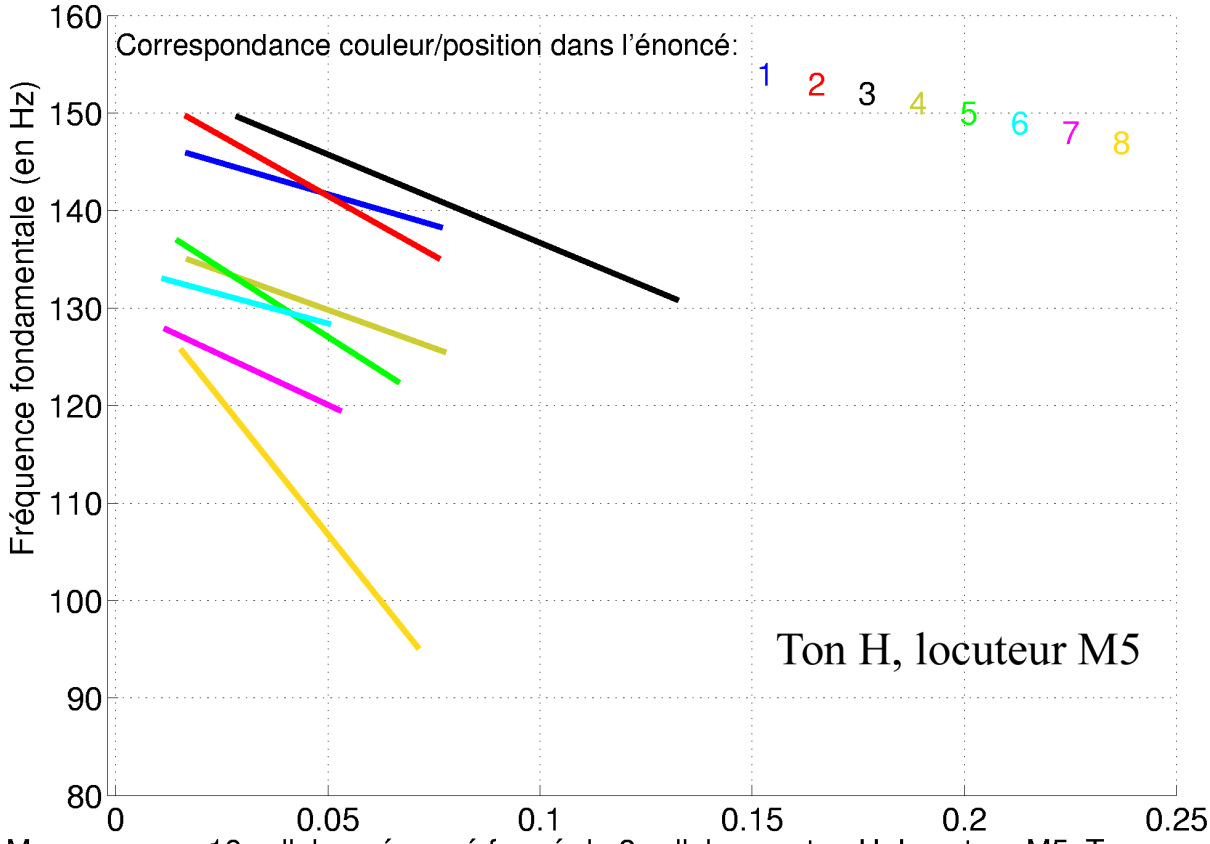


Fig. 2.22. Moyennes sur 16 syllabes; énoncé formé de 8 syllabes au ton H. Locuteur M5. Temps en ms.

Fig. 2.13 à 2.28 : courbes moyennes de F0 et quotient ouvert d'énoncés dont toutes les syllabes portent le même ton (H, M, L ou LH). Nombres impairs : courbes moyennées, nombres pairs : courbes stylisées en deux points.

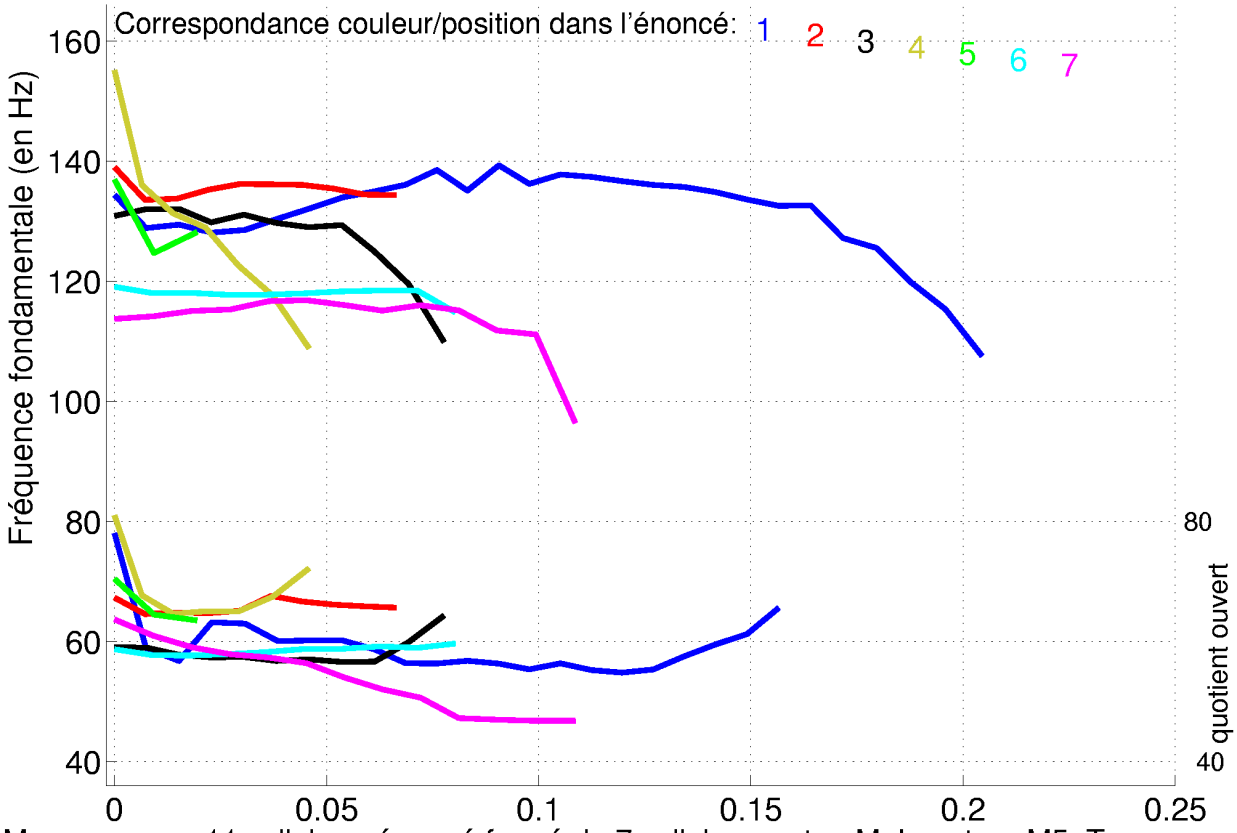


Fig. 2.23. Moyennes sur 11 syllabes; énoncé formé de 7 syllabes au ton M. Locuteur M5. Temps en ms.

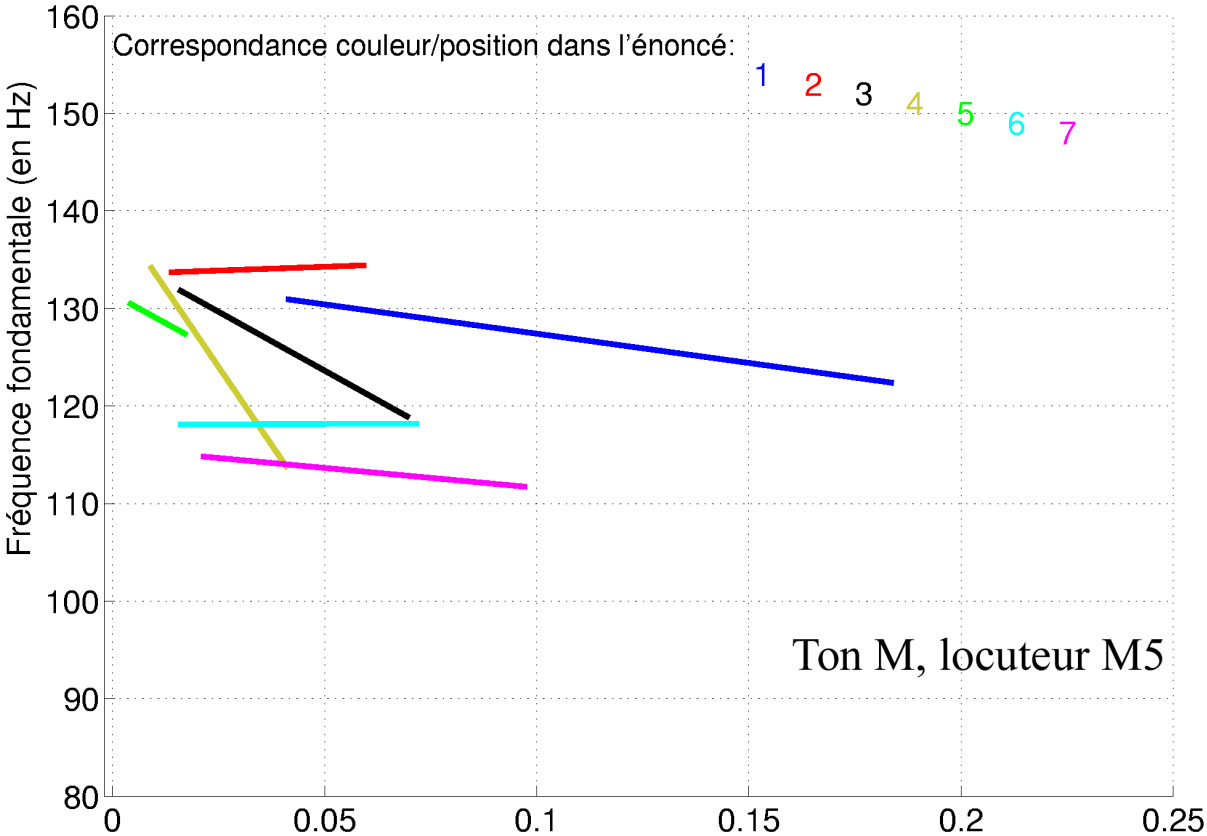


Fig. 2.24. Moyennes sur 11 syllabes; énoncé formé de 7 syllabes au ton M. Locuteur M5. Temps en ms.

Fig. 2.13 à 2.28 : courbes moyennes de F0 et quotient ouvert d'énoncés dont toutes les syllabes portent le même ton (H, M, L ou LH). Nombres impairs : courbes moyennées, nombres pairs : courbes stylisées en deux points.

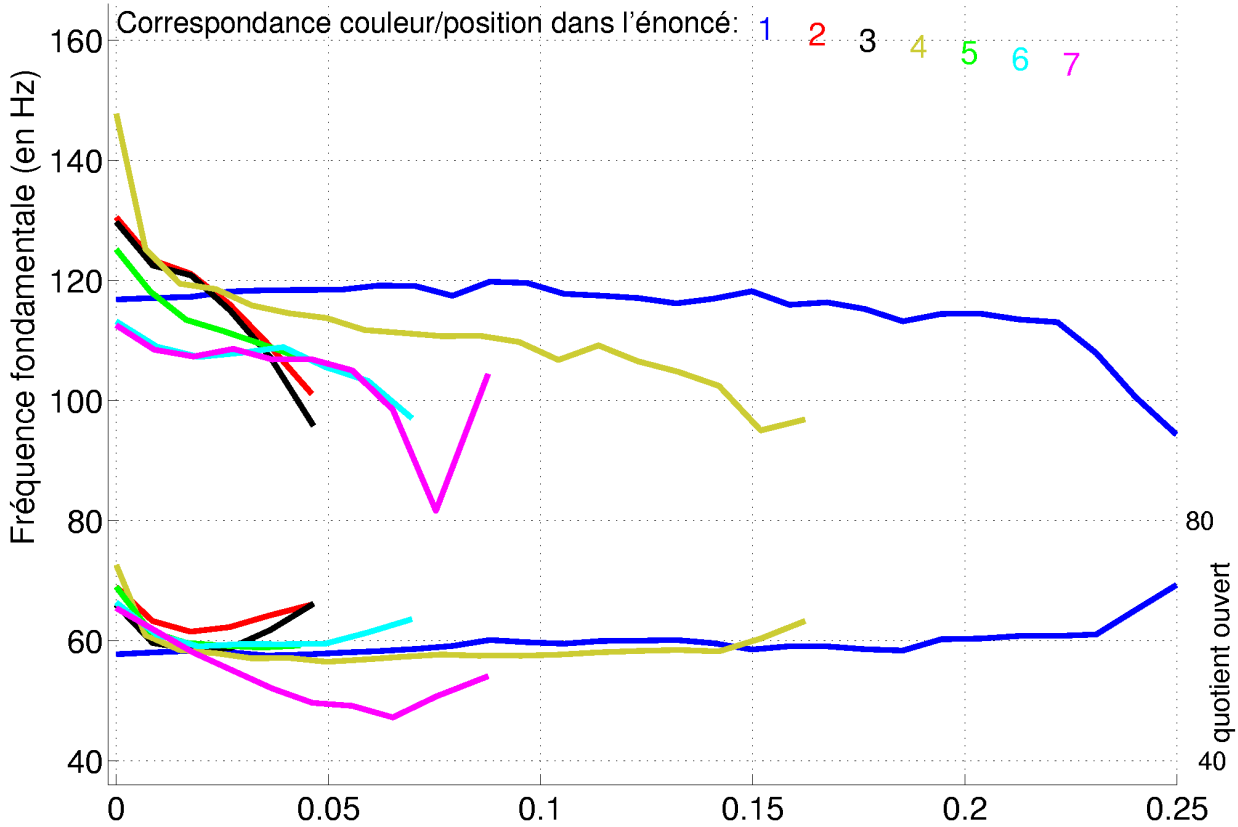


Fig. 2.25. Moyennes sur 14 syllabes; énoncé formé de 7 syllabes au ton L. Locuteur M5. Temps en ms.

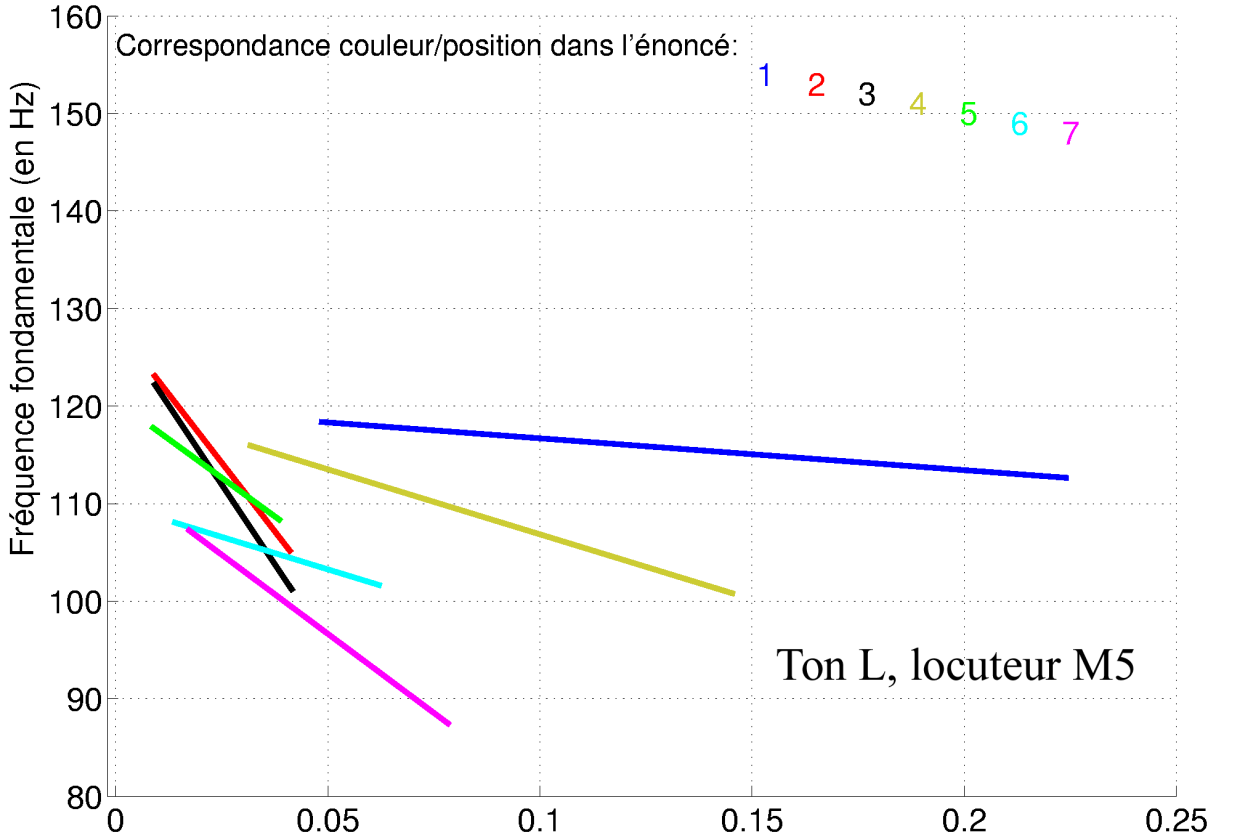


Fig. 2.26. Moyennes sur 14 syllabes; énoncé formé de 7 syllabes au ton L. Locuteur M5. Temps en ms.

Fig. 2.13 à 2.28 : courbes moyennes de F0 et quotient ouvert d'énoncés dont toutes les syllabes portent le même ton (H, M, L ou LH). Nombres impairs : courbes moyennées, nombres pairs : courbes stylisées en deux points.

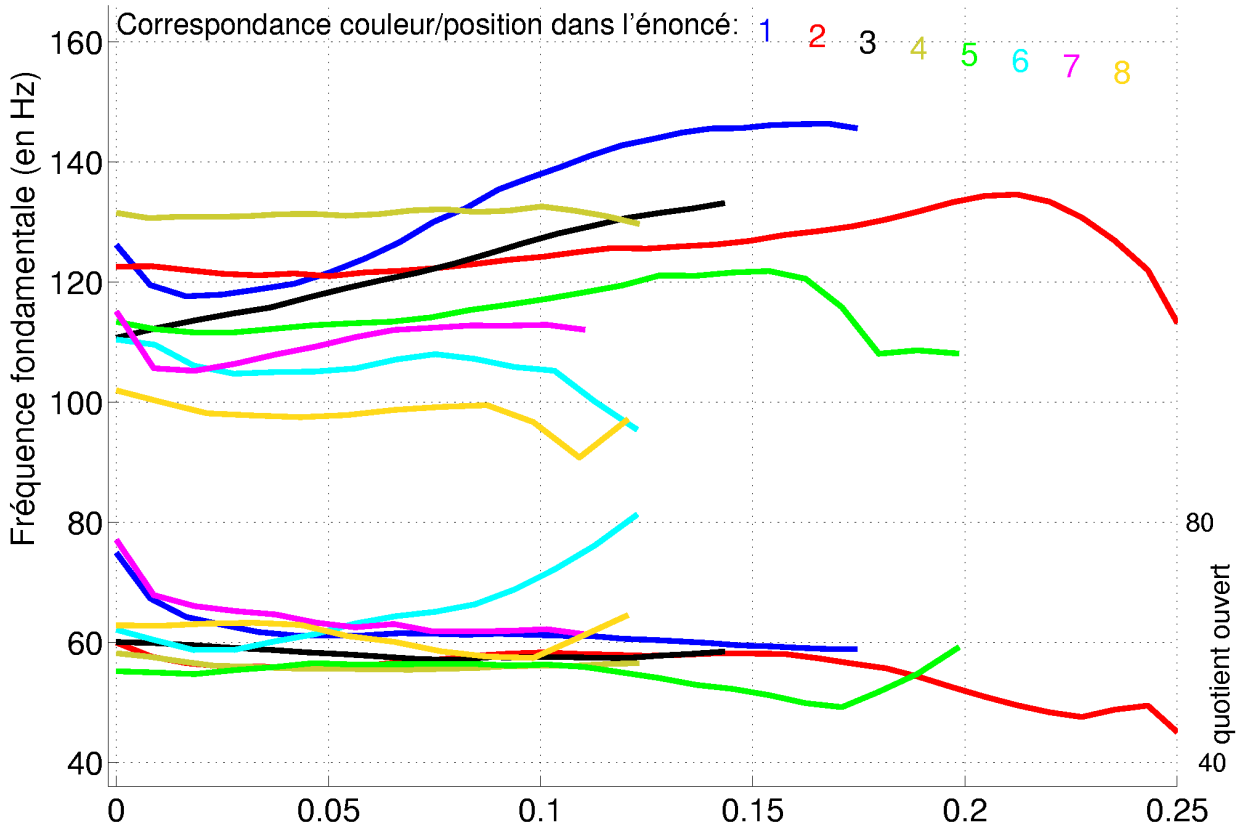


Fig. 2.27. Moyennes sur 39 syllabes; énoncé formé de 8 syllabes au ton R. Locuteur M5. Temps en ms.

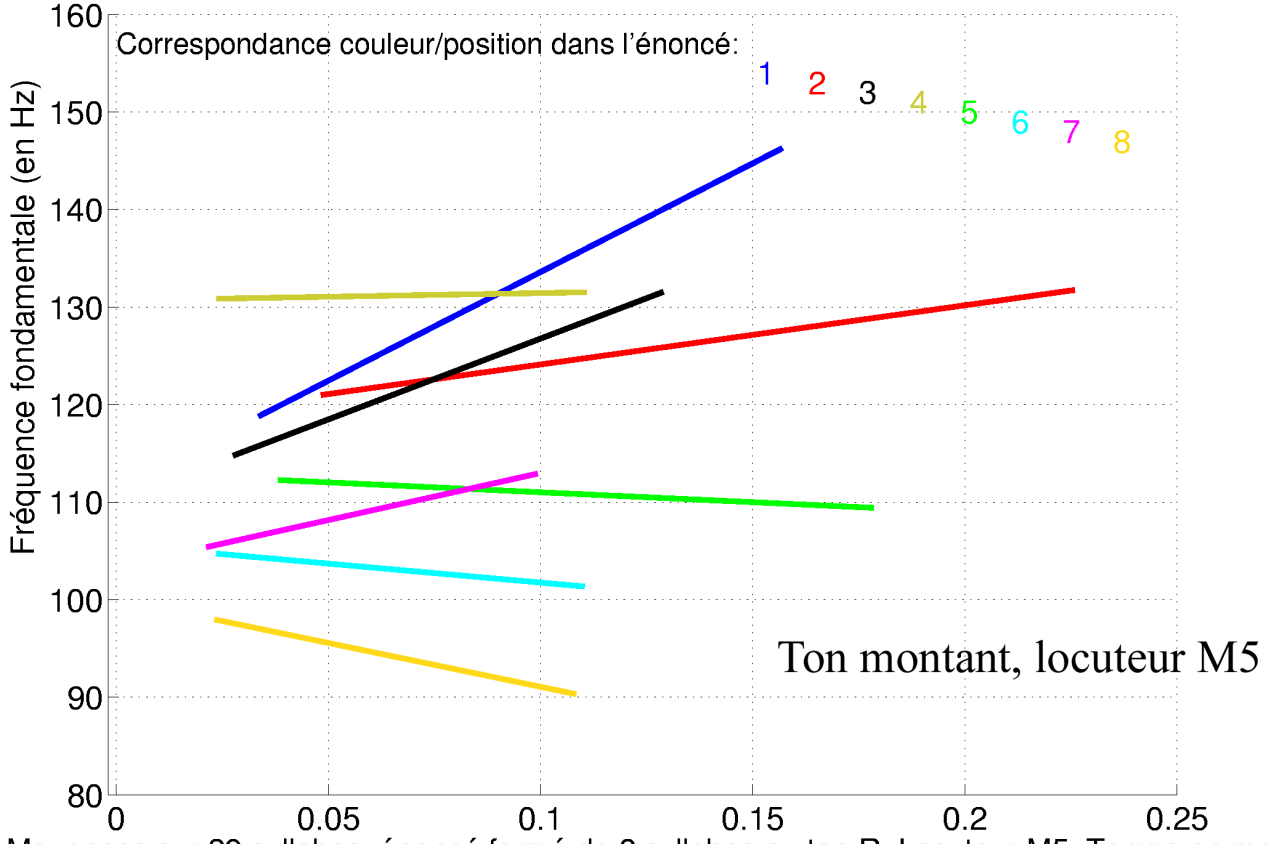


Fig. 2.28. Moyennes sur 39 syllabes; énoncé formé de 8 syllabes au ton R. Locuteur M5. Temps en ms.

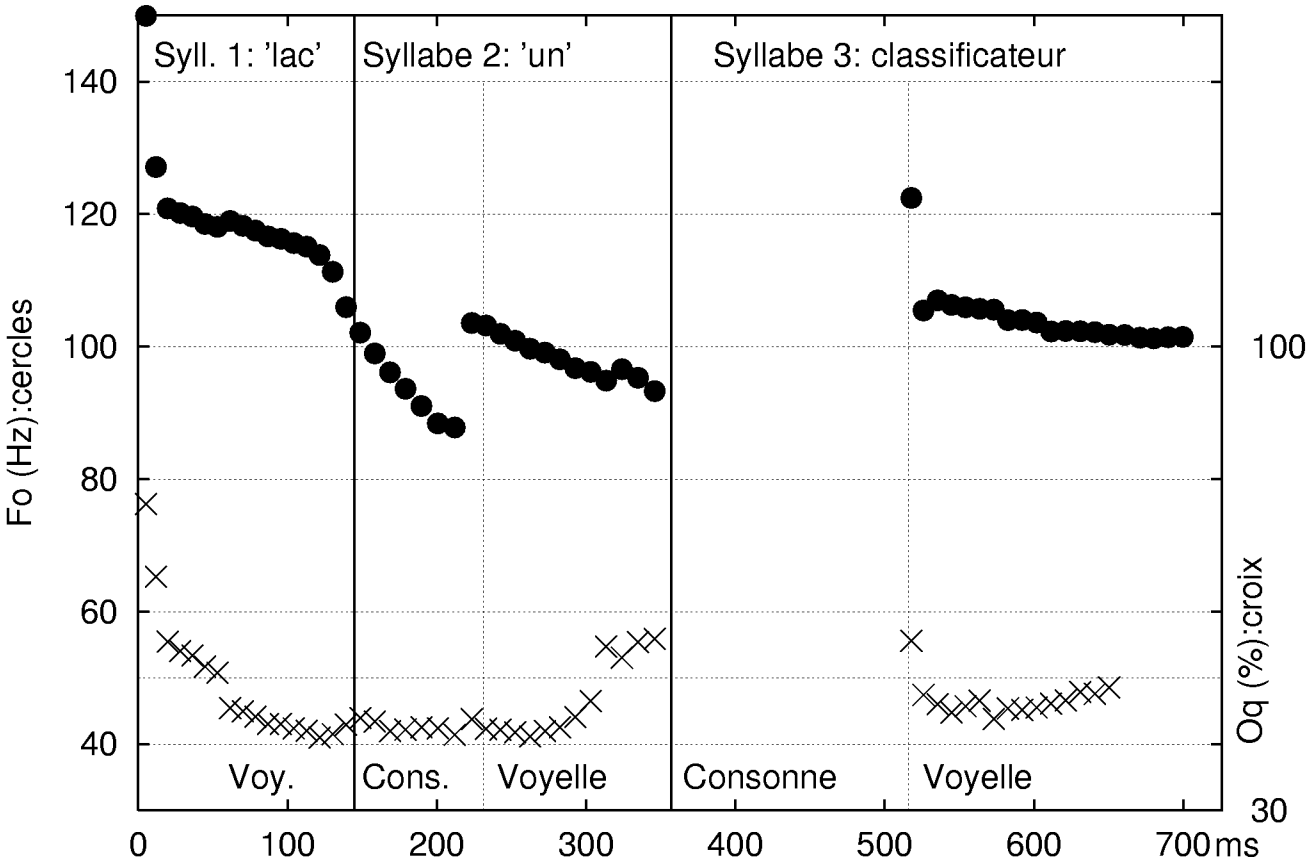


Fig. 2.29. Courbes de F0 et de quotient ouvert sur l'expression " un lac ". Locuteur M1.

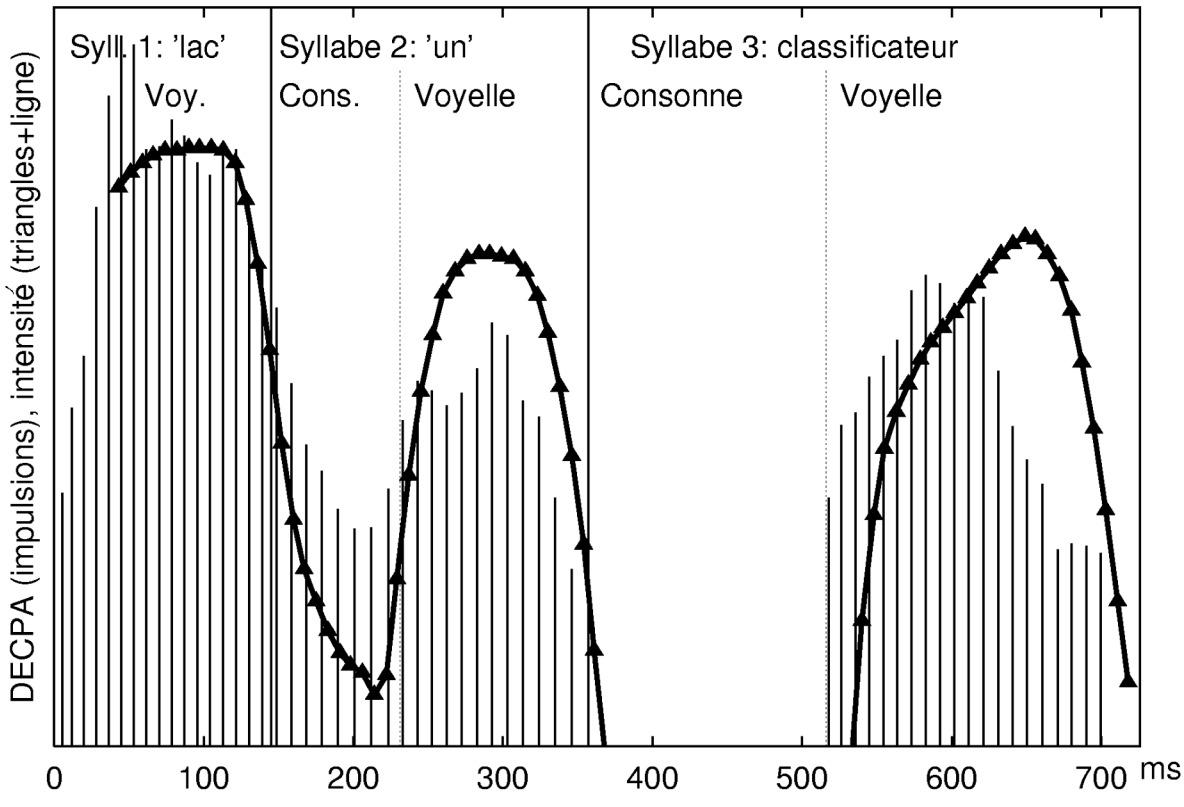


Fig. 2.30. Courbes de DECPA et d'intensité sur l'expression " un lac ". Locuteur M1.

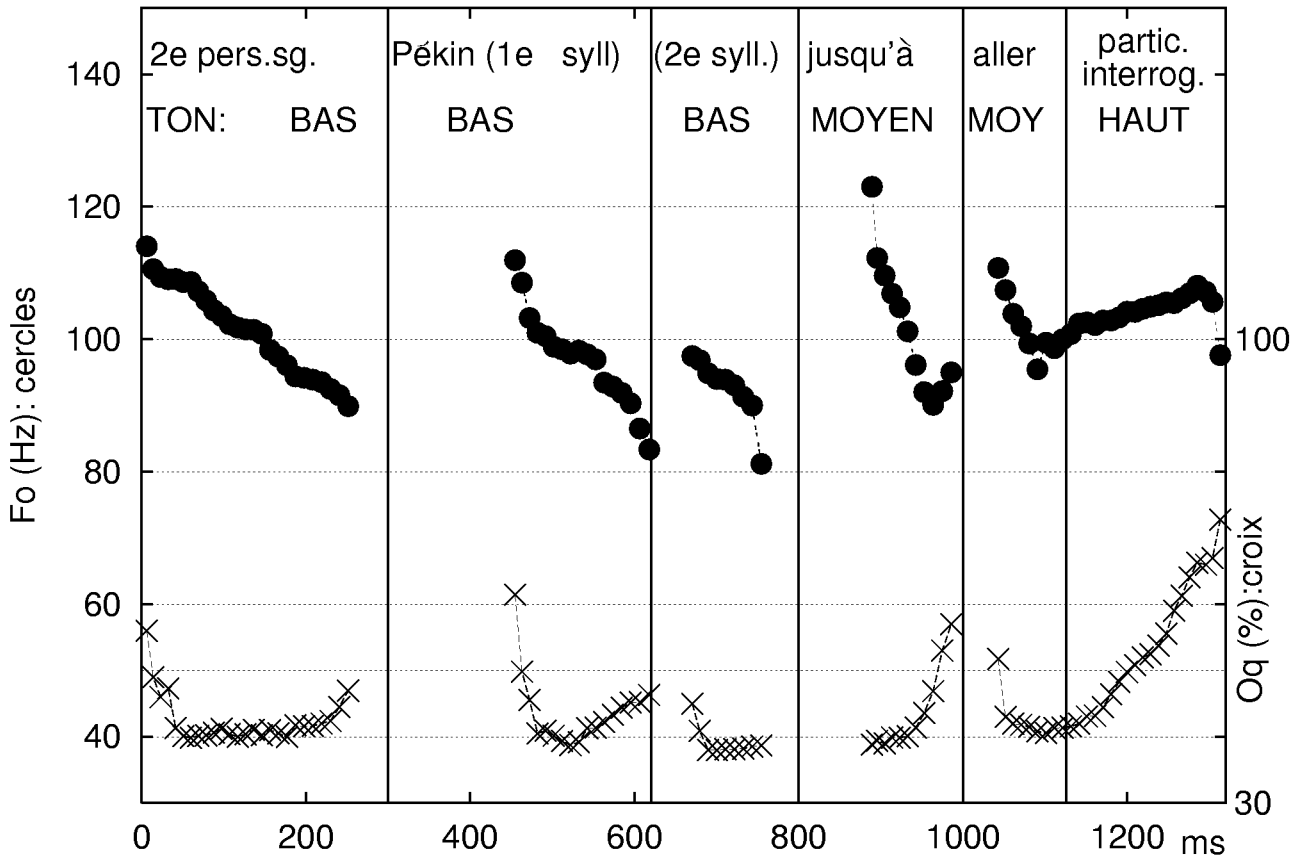


Fig. 2.31. Courbes de F0 et quotient ouvert de la phrase " Je vais à Pékin ". Locuteur M1.

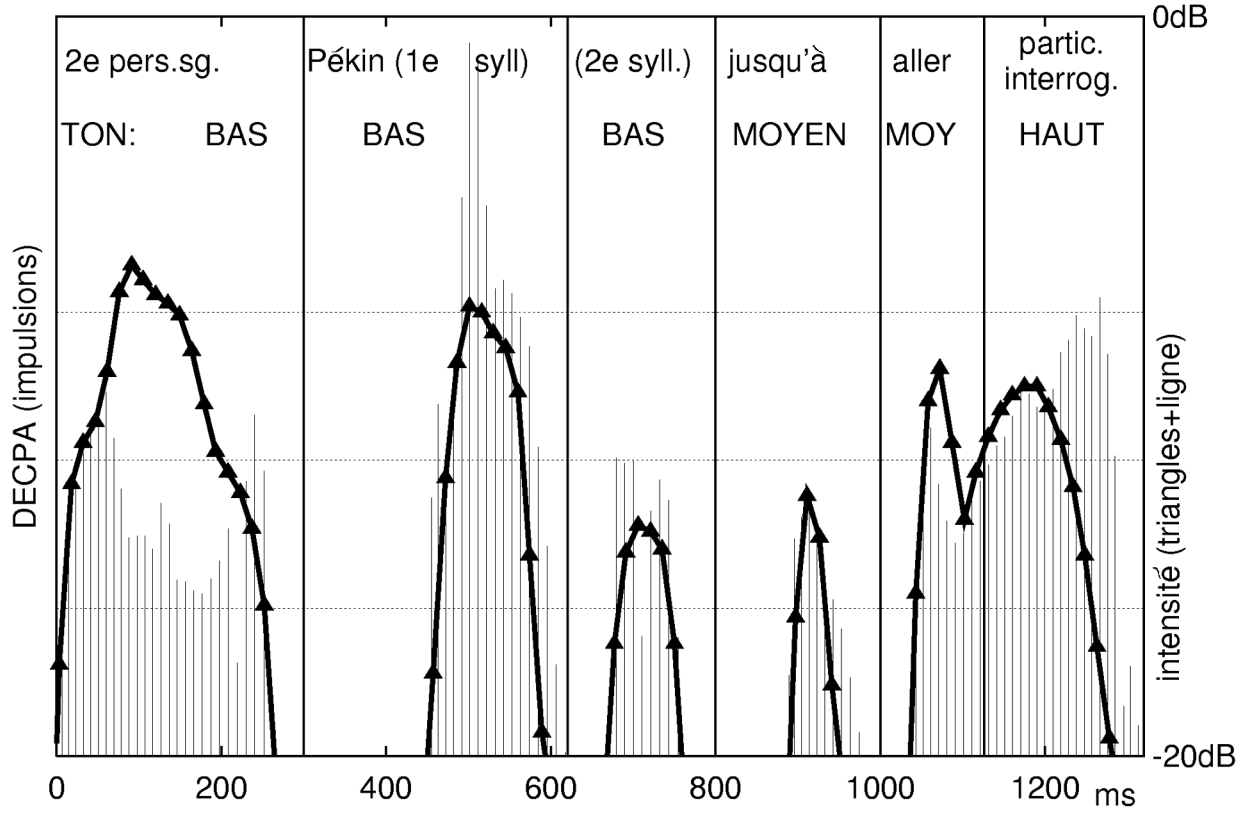


Fig. 2.32. Courbes de DECPA et d'intensité de la phrase " Je vais à Pékin ". Locuteur M1.

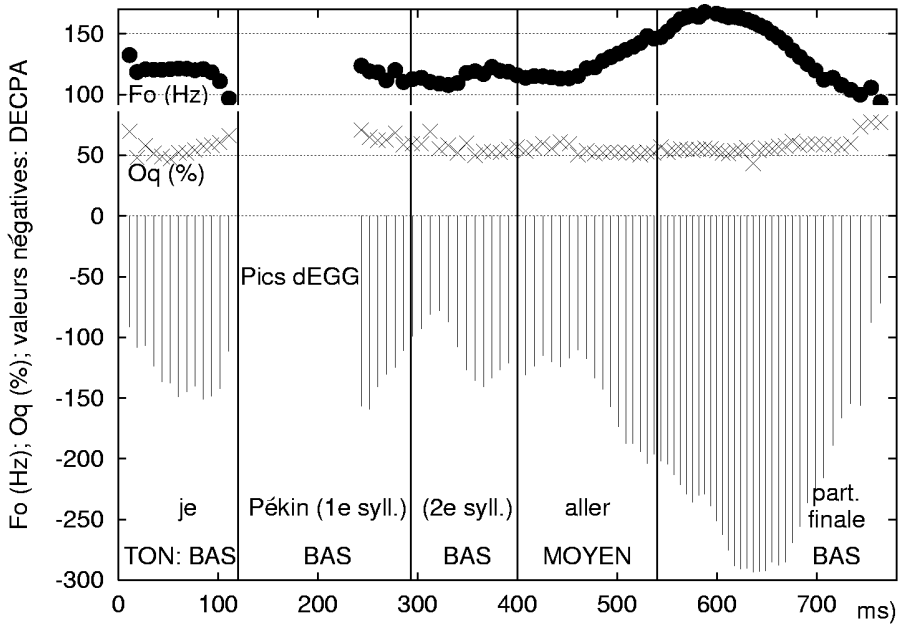


Fig. 2.33. Courbes de F0, quotient ouvert et DECPA de la phrase 24 (" C'est à Pékin que je vais "). Locuteur M1.

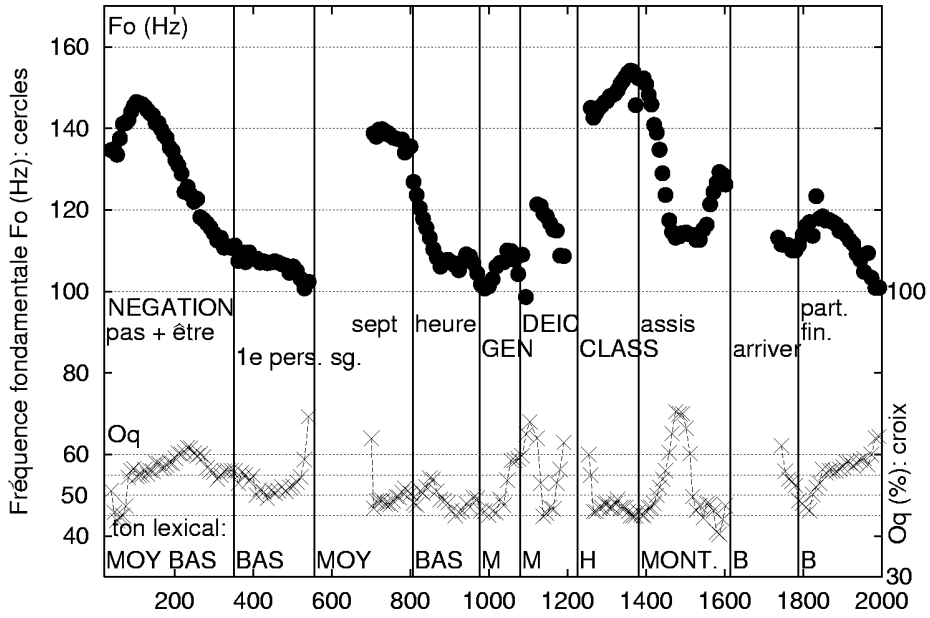


Fig. 2.34. Courbes de F0 et quotient ouvert de la phrase 17' (" Non, c'est par... "). Locuteur M1.

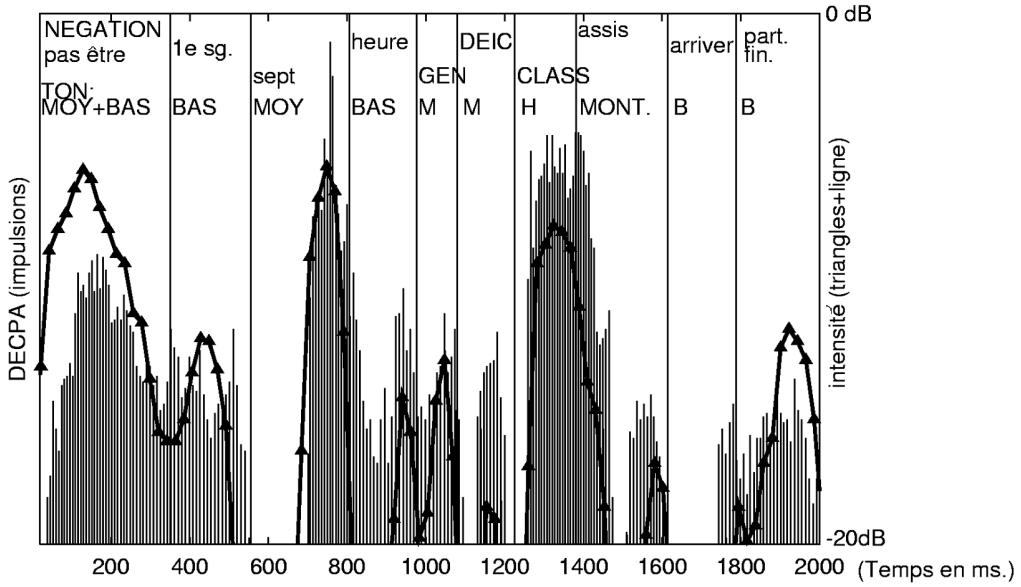


Fig. 2.35. Courbes de DECPA et d'intensité de la phrase 17' (" Non, c'est par... "). Locuteur M1.

Fig. 3.1 à 3.12 : courbes moyennes de F0 et quotient ouvert des mots-cible en anglais, permettant une comparaison entre conditions de lecture.
Monosyllabes (figures à numéros impairs) et disyllabes (figures à numéros pairs), locuteurs M2 à M7.

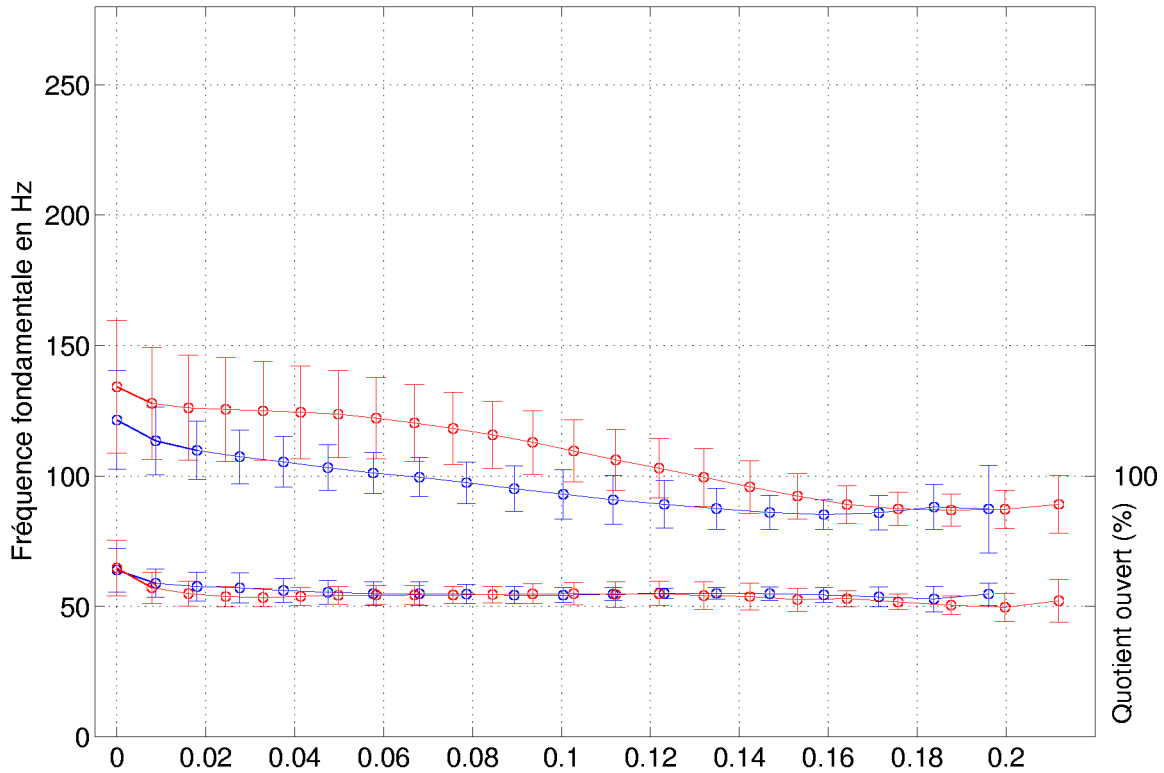


Fig. 3.1. Monosyllabes, cond. S (en bleu) et I (en rouge). Locuteur M2. 48 syll. Temps en ms.

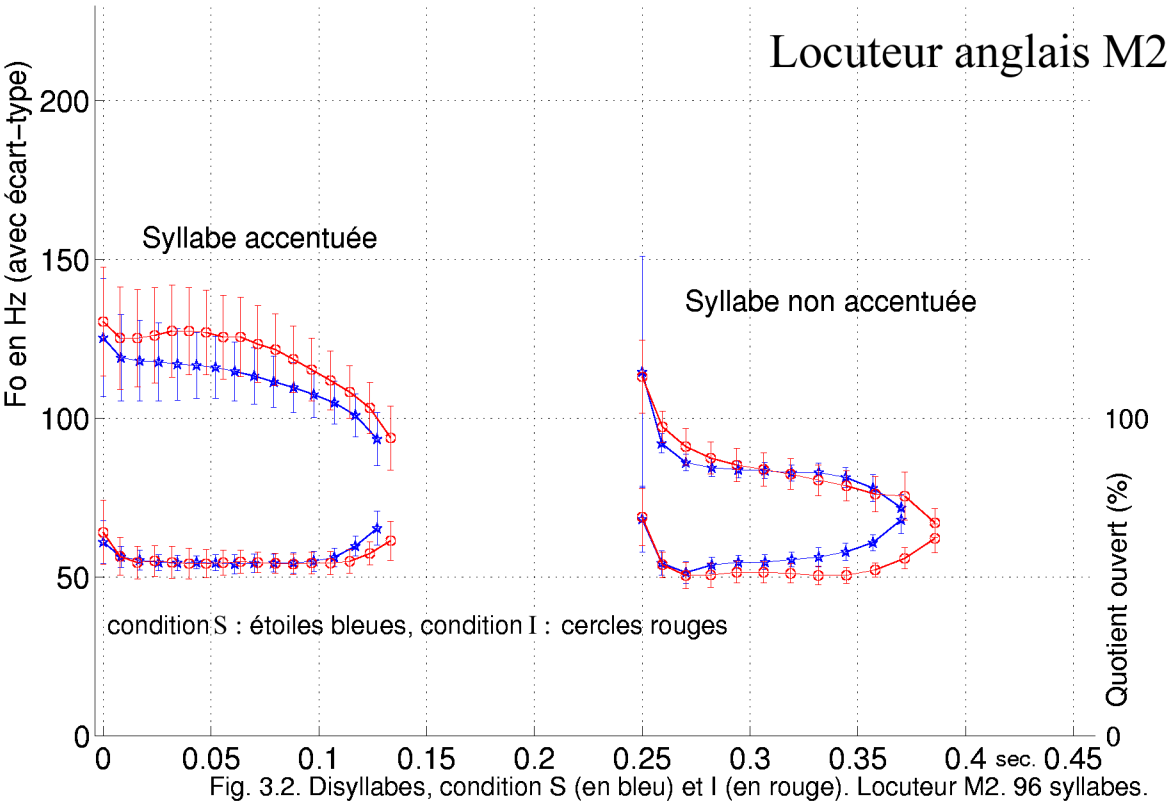


Fig. 3.2. Disyllabes, condition S (en bleu) et I (en rouge). Locuteur M2. 96 syllabes.

Fig. 3.1 à 3.12 : courbes moyennes de F0 et quotient ouvert des mots-cible en anglais, permettant une comparaison entre conditions de lecture.
Monosyllabes (figures à numéros impairs) et disyllabes (figures à numéros pairs), locuteurs M2 à M7.

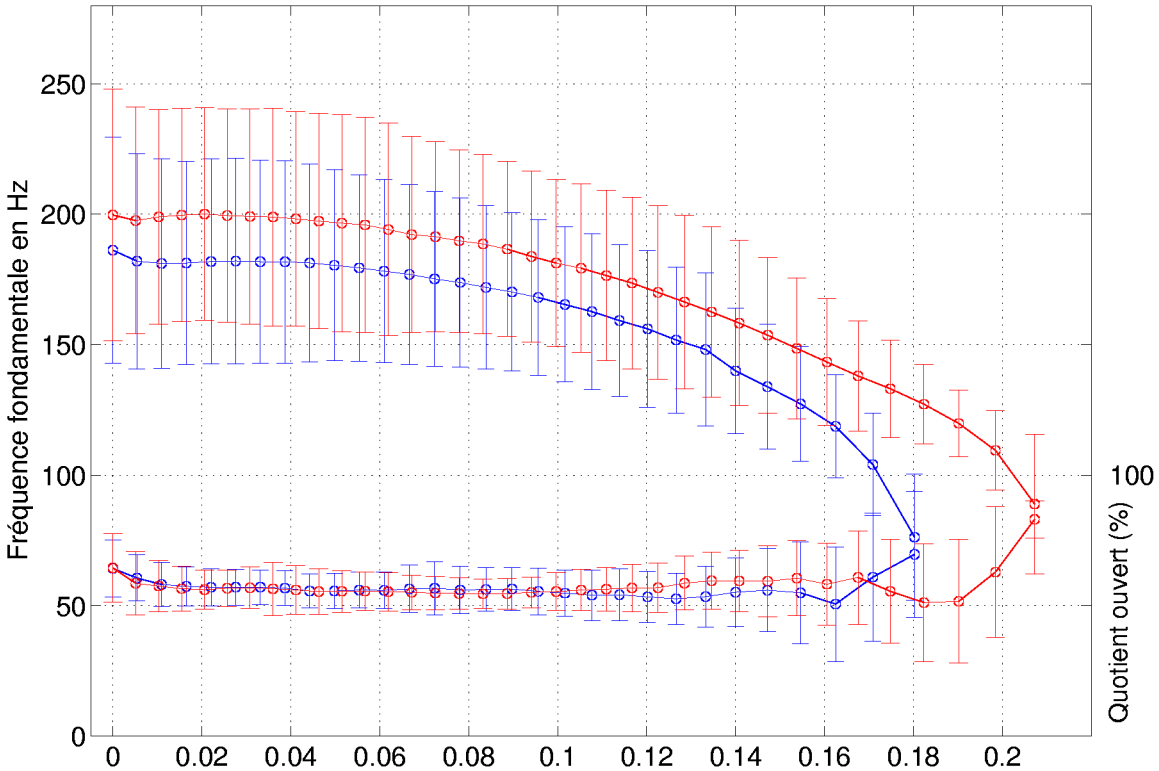


Fig. 3.3. Monosyllabes, cond. S (en bleu) et I (en rouge). Locuteur M3. 48 syll. Temps en ms.

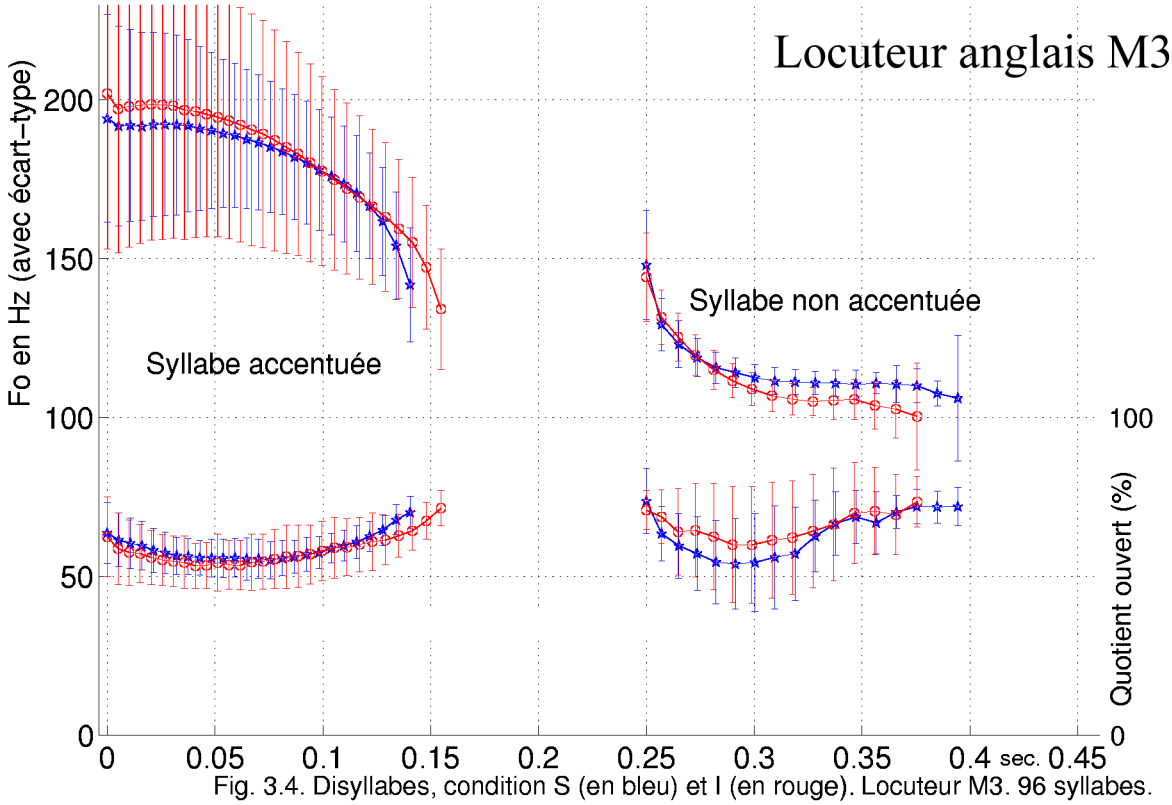


Fig. 3.4. Disyllabes, condition S (en bleu) et I (en rouge). Locuteur M3. 96 syllabes.

Fig. 3.1 à 3.12 : courbes moyennes de F0 et quotient ouvert des mots-cible en anglais, permettant une comparaison entre conditions de lecture.
Monosyllabes (figures à numéros impairs) et disyllabes (figures à numéros pairs), locuteurs M2 à M7.

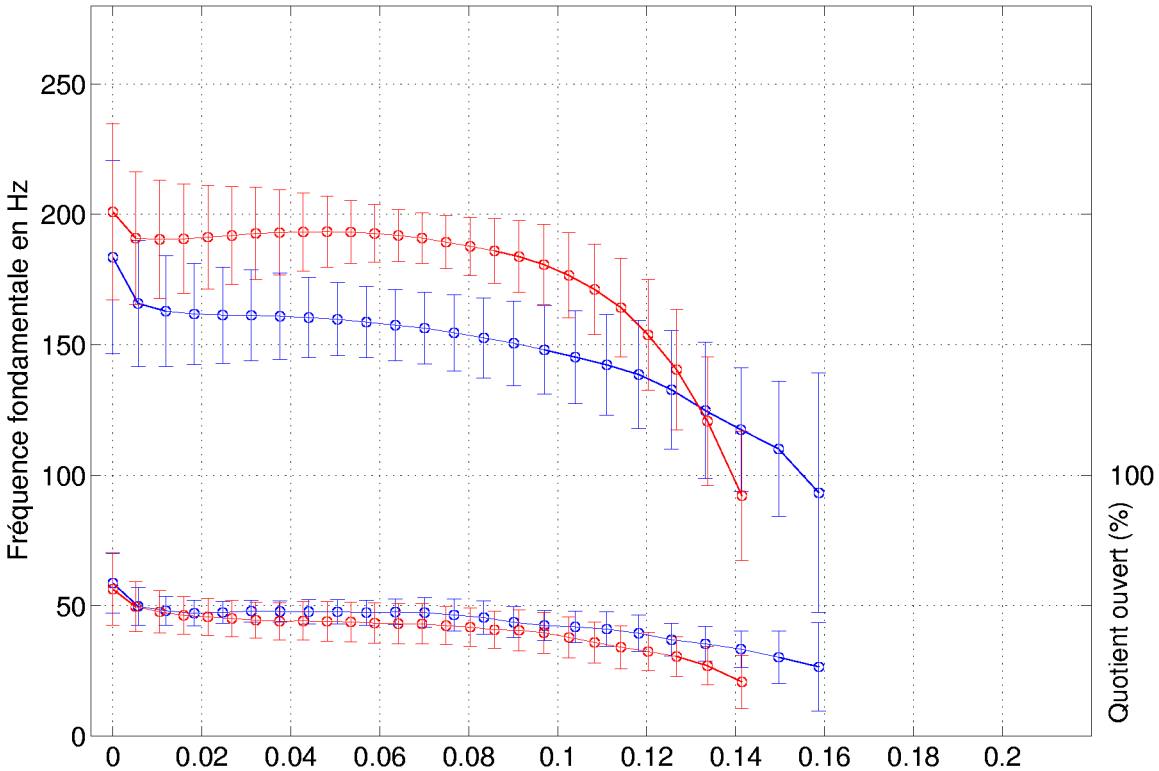


Fig. 3.5. Monosyllabes, cond. S (en bleu) et I (en rouge). Locuteur M4. 48 syll. Temps en ms.

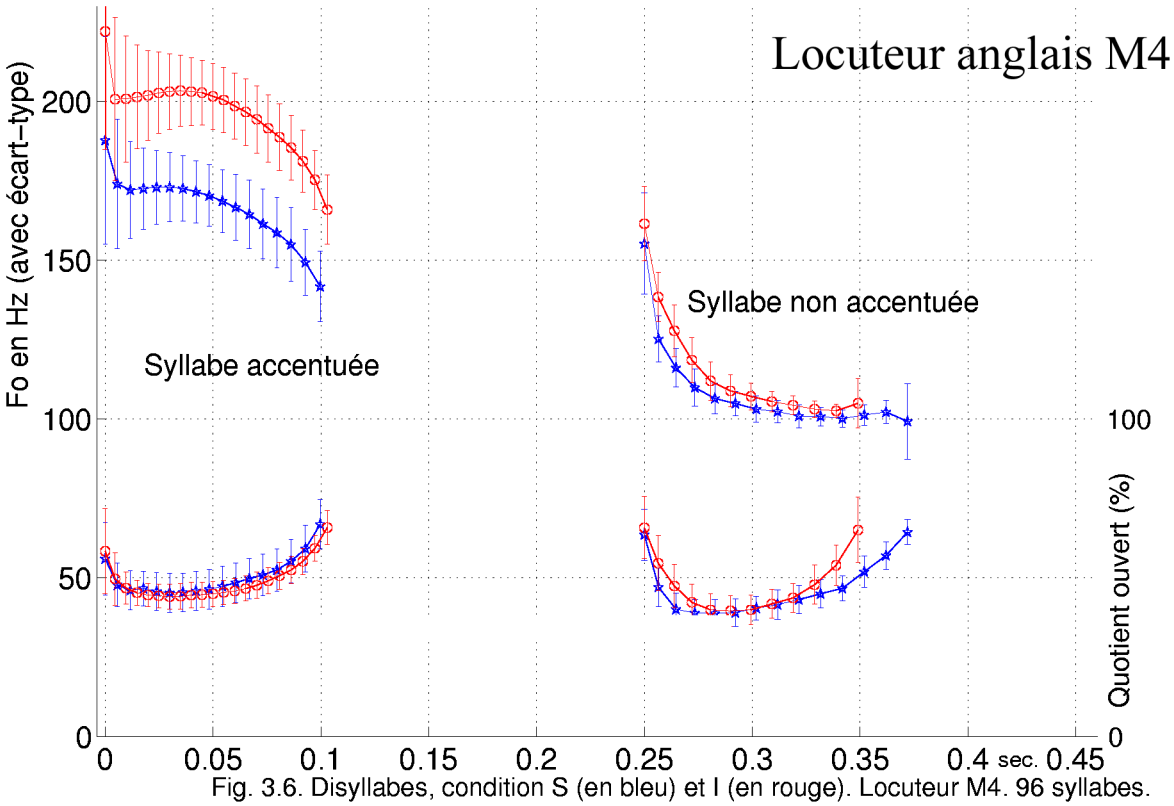


Fig. 3.6. Disyllabes, condition S (en bleu) et I (en rouge). Locuteur M4. 96 syllabes.

Fig. 3.1 à 3.12 : courbes moyennes de F0 et quotient ouvert des mots-cible en anglais, permettant une comparaison entre conditions de lecture.
Monosyllabes (figures à numéros impairs) et disyllabes (figures à numéros pairs), locuteurs M2 à M7.

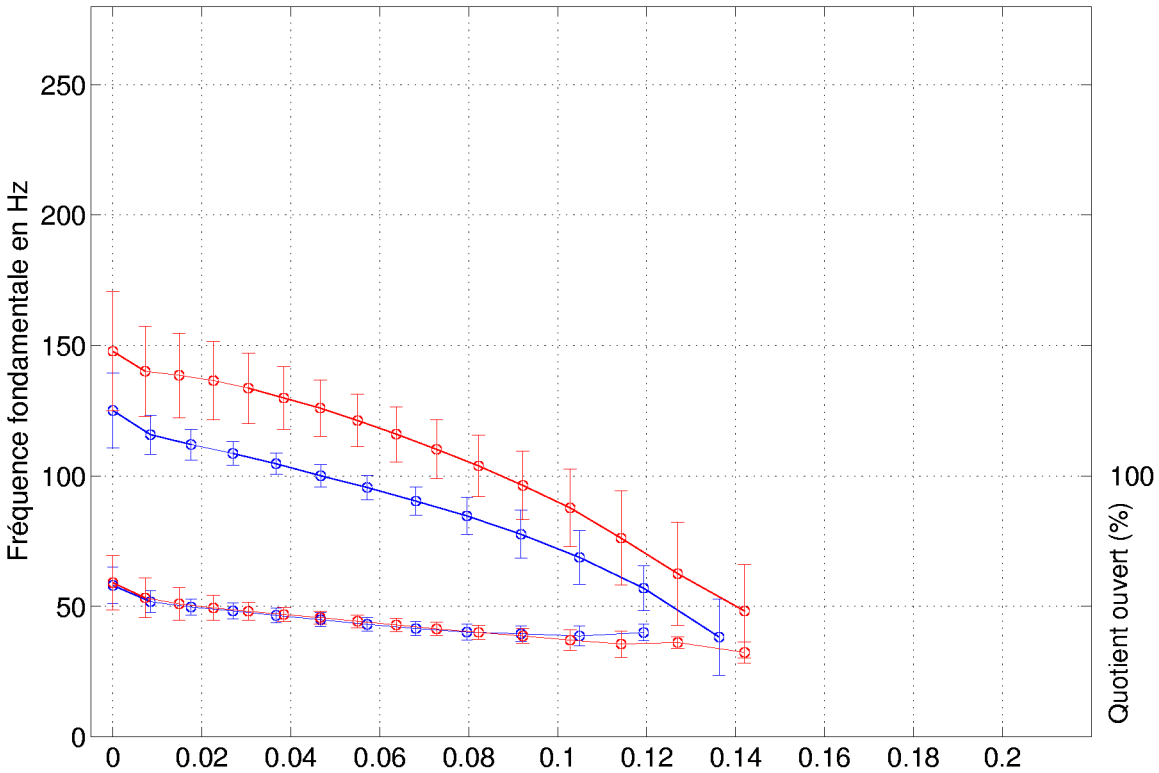


Fig. 3.7. Monosyllabes, cond. C (en bleu) et E (en rouge). Locuteur M5. 48 syll. Temps en ms.

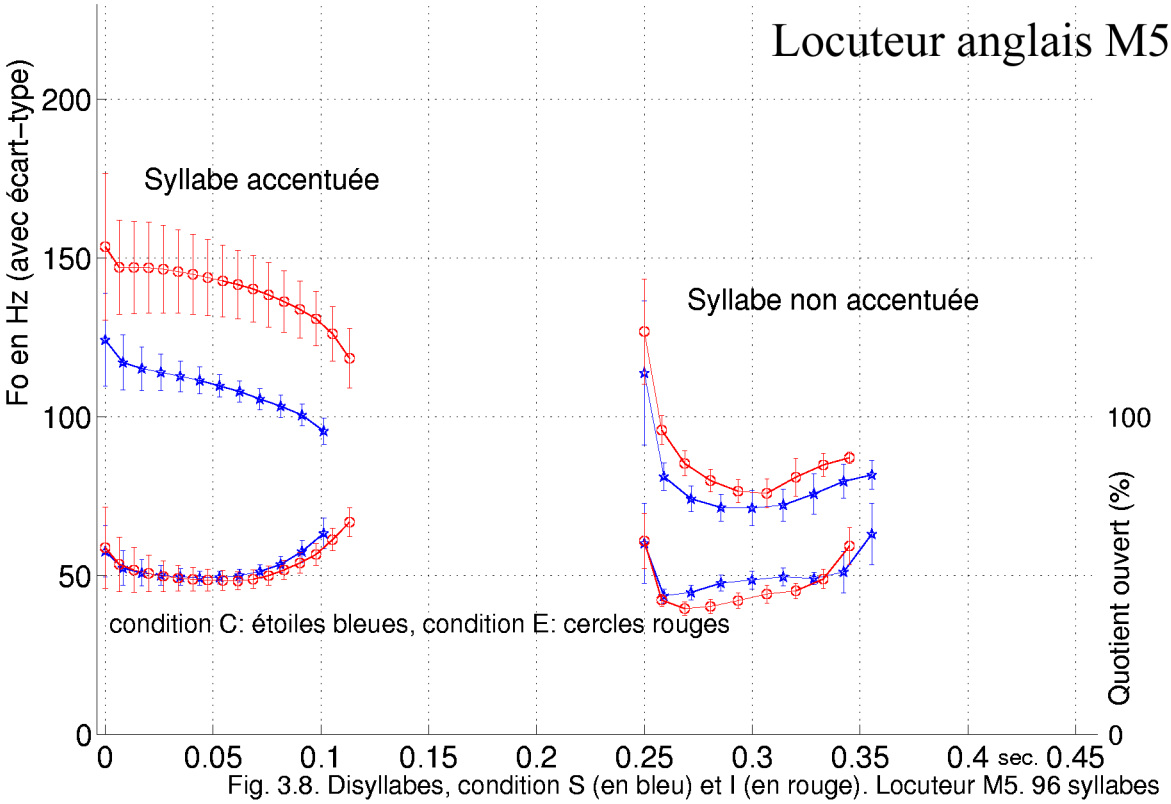


Fig. 3.8. Disyllabes, condition S (en bleu) et I (en rouge). Locuteur M5. 96 syllabes

Fig. 3.1 à 3.12 : courbes moyennes de F0 et quotient ouvert des mots-cible en anglais, permettant une comparaison entre conditions de lecture.
Monosyllabes (figures à numéros impairs) et disyllabes (figures à numéros pairs), locuteurs M2 à M7.

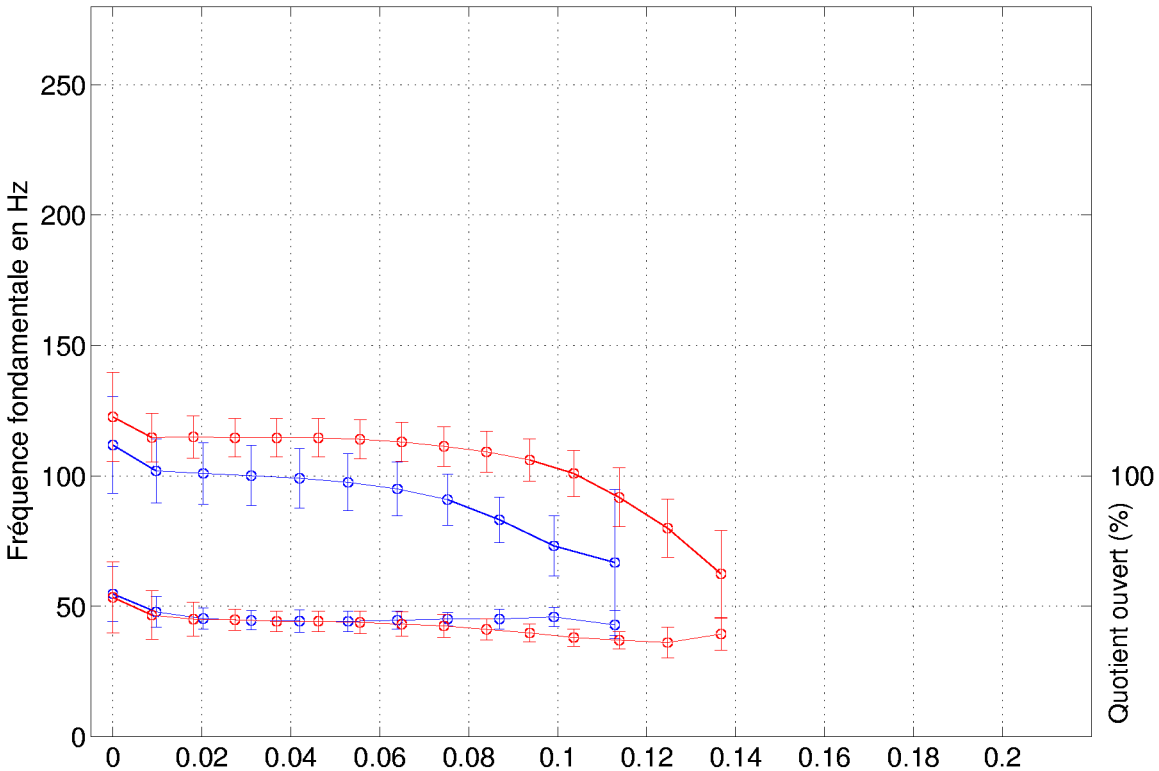


Fig. 3.9. Monosyllabes, cond. S (en bleu) et I (en rouge). Locuteur M6. 48 syll. Temps en ms.

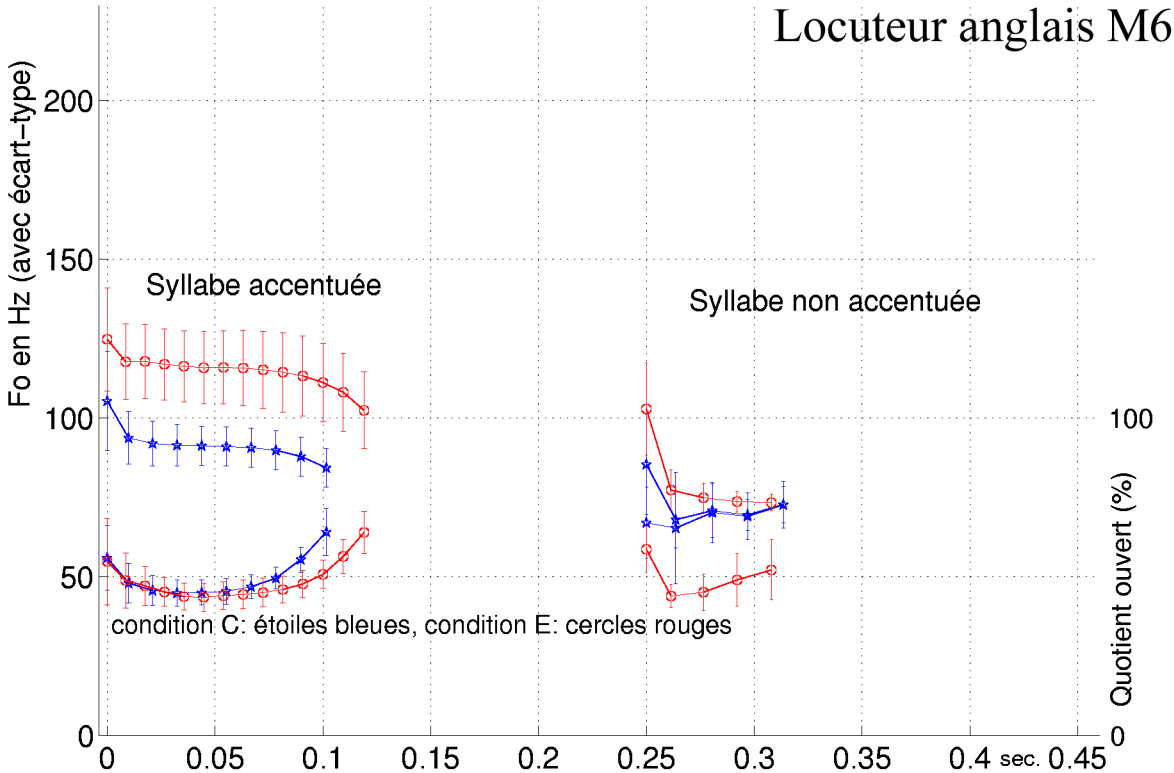


Fig. 3.10. Disyllabes, condition S (en bleu) et I (en rouge). Locuteur M6. 96 syllabes.

Fig. 3.1 à 3.12 : courbes moyennes de F0 et quotient ouvert des mots-cible en anglais, permettant une comparaison entre conditions de lecture.
Monosyllabes (figures à numéros impairs) et disyllabes (figures à numéros pairs), locuteurs M2 à M7.

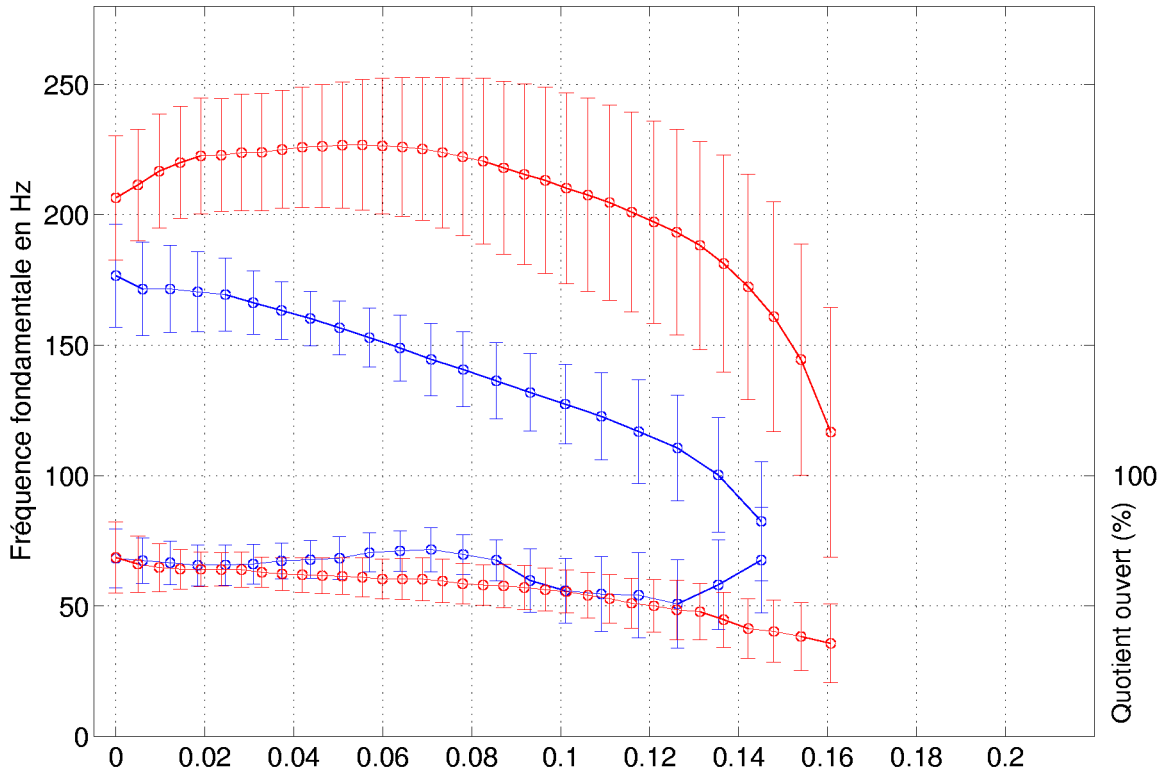


Fig. 3.11. Monosyllabes, cond. S (en bleu) et I (en rouge). Locuteur M7. 48 syll. Temps en ms.

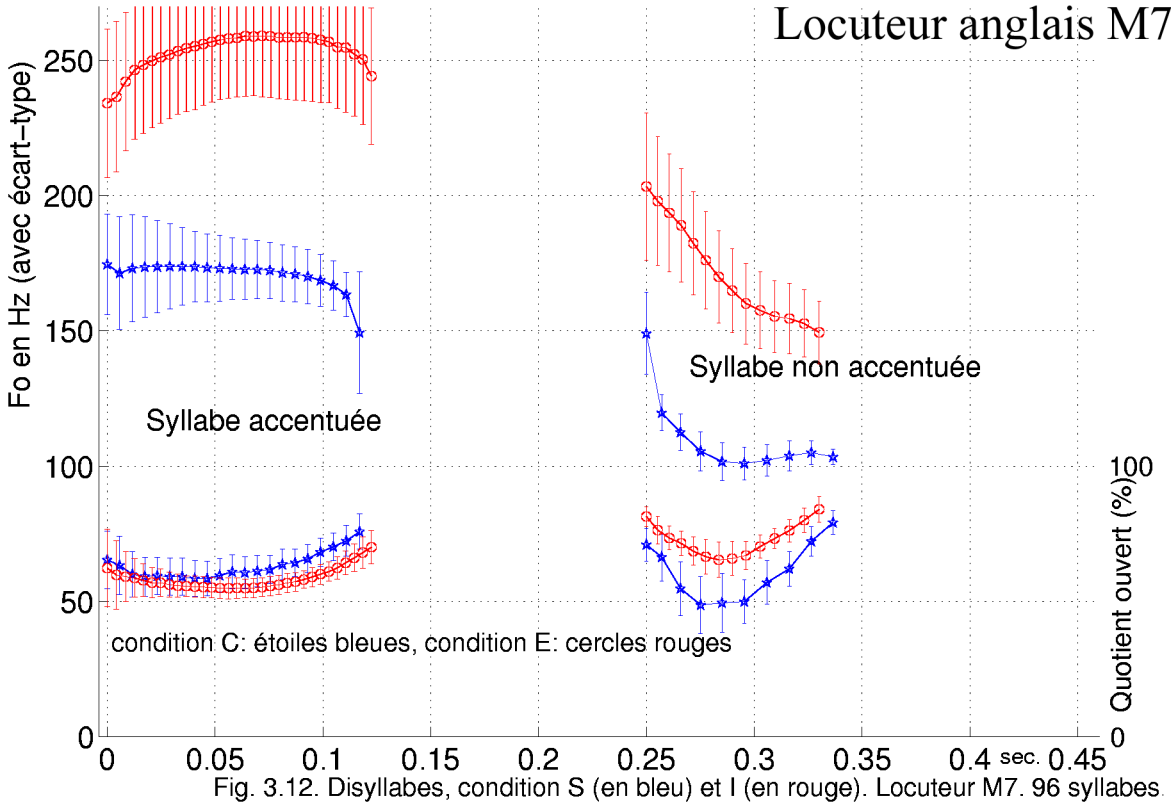


Fig. 3.12. Disyllabes, condition S (en bleu) et I (en rouge). Locuteur M7. 96 syllabes

Fig. 3.13 à 3.27 : estimation des trois premiers formants des voyelles des monosyllabes en anglais. Résultats bruts (figures 3.13a, 3.14a...) et après suppression de valeurs aberrantes suite à vérification visuelle (figures 3.13b, 3.14b...), locuteurs M3 à M7.

Voyelle : / ɑː /

Colonne de gauche : résultats bruts du logiciel Praat

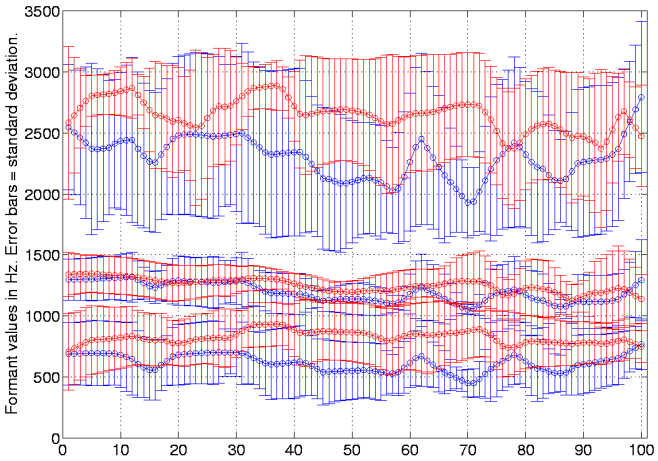


Fig. 3.13a. Condition de lecture S : en bleu, I : en rouge.

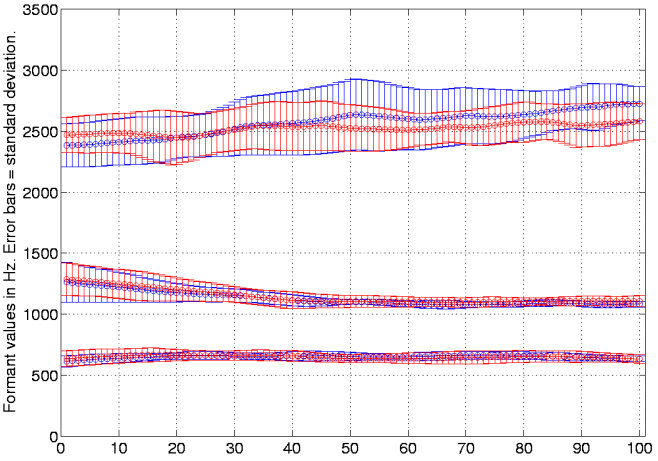


Fig. 3.14a. Condition de lecture S : en bleu, I : en rouge.

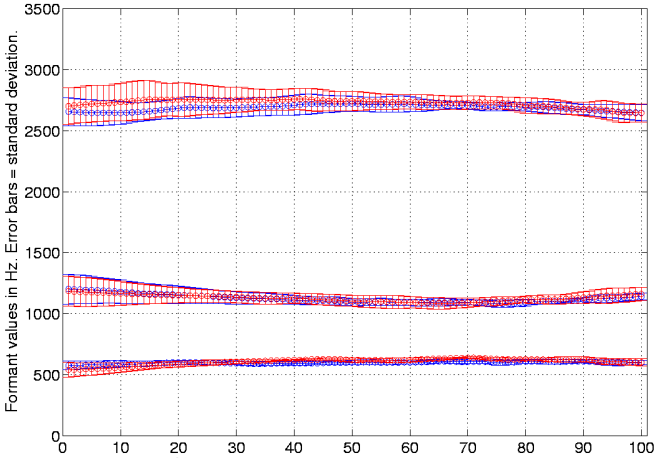


Fig. 3.15a. Condition de lecture S : en bleu, I : en rouge.

Colonne de droite : résultats vérifiés visuellement

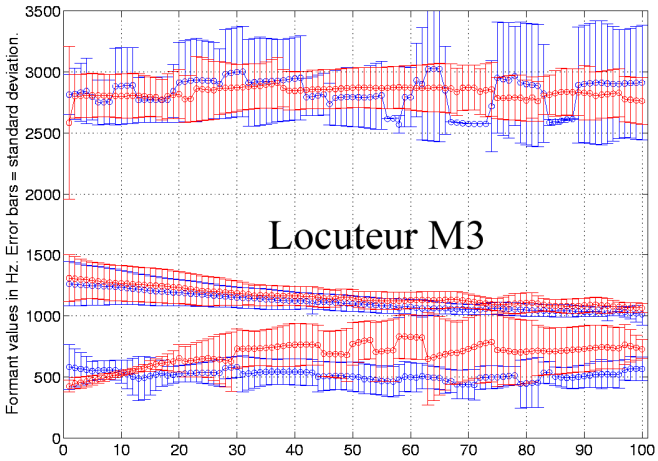


Fig. 3.13b. Condition de lecture S : en bleu, I : en rouge.

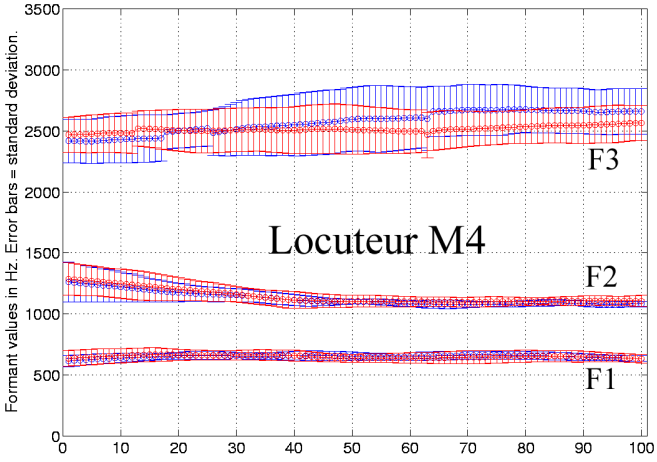


Fig. 3.14b. Condition de lecture S : en bleu, I : en rouge.

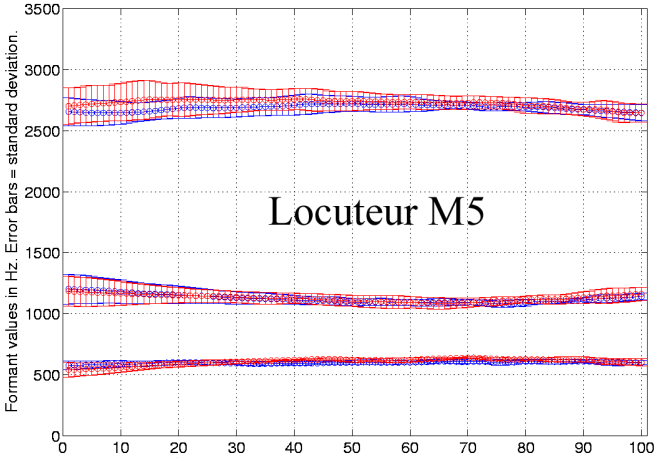


Fig. 3.15b. Condition de lecture S : en bleu, I : en rouge.

Fig. 3.13 à 3.27 : estimation des trois premiers formants des voyelles des monosyllabes en anglais.
Résultats bruts (figures 3.16a, 3.17a...) et après suppression de valeurs aberrantes suite à vérification visuelle (figures 3.16b, 3.17b...), locuteurs M3 à M7.

Colonne de gauche :
résultats bruts du logiciel Praat

Colonne de droite :
résultats vérifiés visuellement

Voyelle : / ɑː /

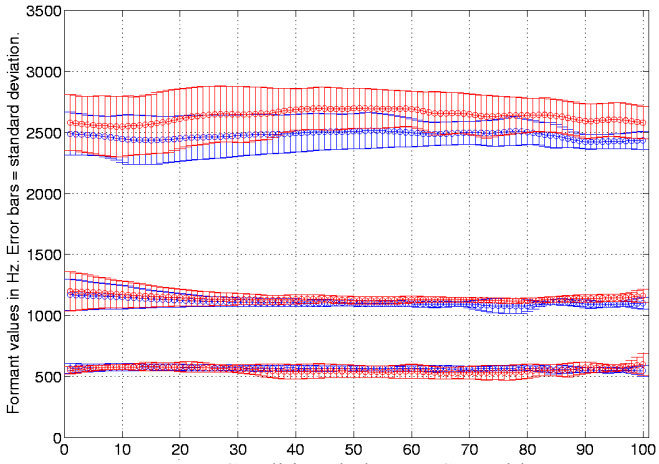


Fig. 3.16a. Condition de lecture S : en bleu, I : en rouge.

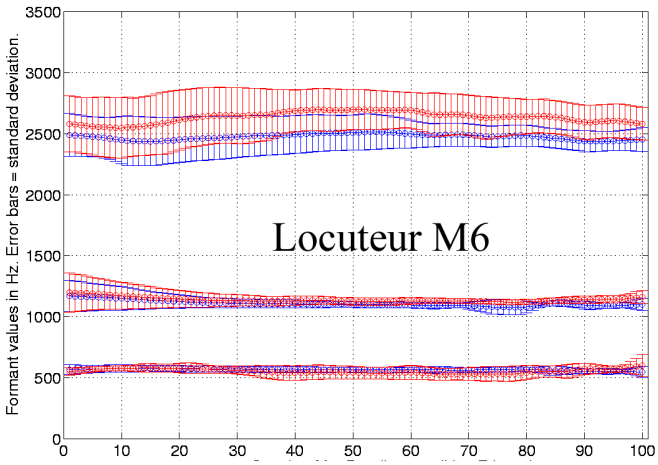


Fig. 3.16b Condition de lecture S : en bleu, I : en rouge.

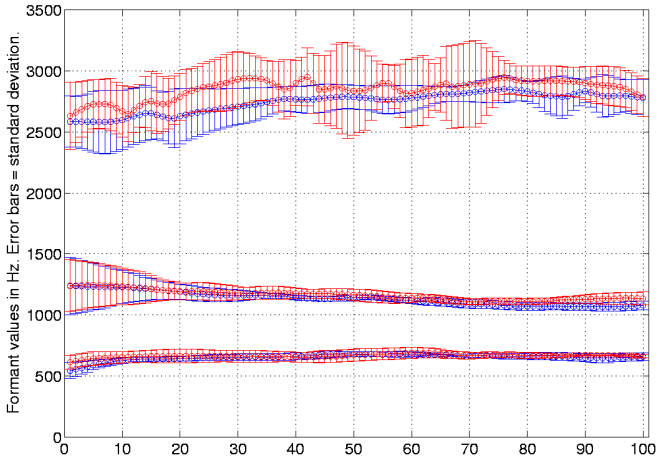


Fig. 3.17a. Condition de lecture S : en bleu, I : en rouge.

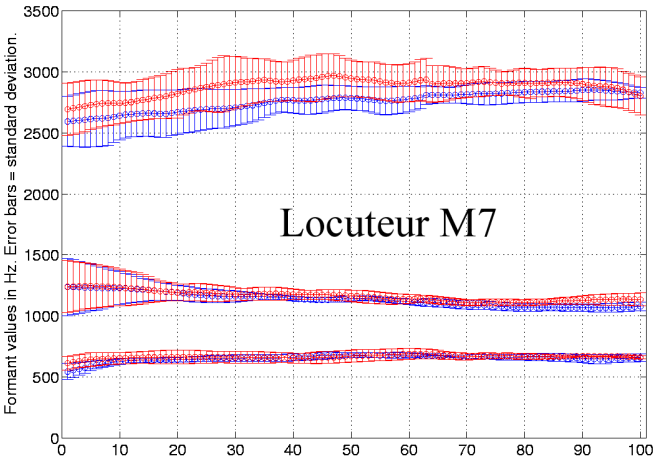


Fig. 3.17b. Condition de lecture S : en bleu, I : en rouge.

Voyelle : / iː /

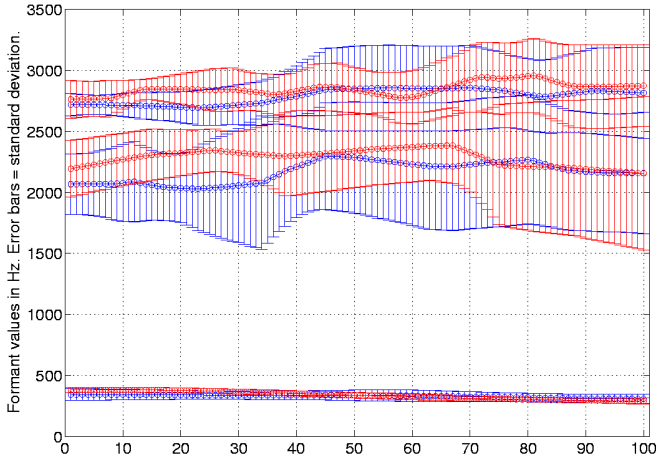


Fig. 3.18a. Condition de lecture S : en bleu, I : en rouge.

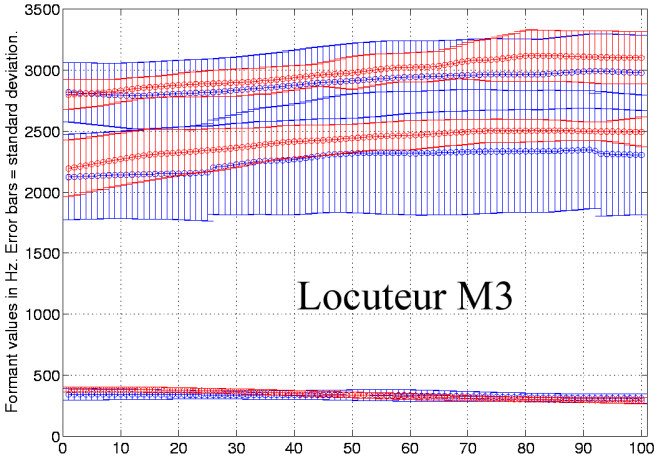


Fig. 3.18b. Condition de lecture S : en bleu, I : en rouge.

Fig. 3.13 à 3.27 : estimation des trois premiers formants des voyelles des monosyllabes en anglais. Résultats bruts (figures 3.19a, 3.20a...) et après suppression de valeurs aberrantes suite à vérification visuelle (figures 3.19b, 3.20b...), locuteurs M3 à M7.

Colonne de gauche :
résultats bruts du logiciel Praat

Colonne de droite :
résultats vérifiés visuellement

Voyelle : / i: /

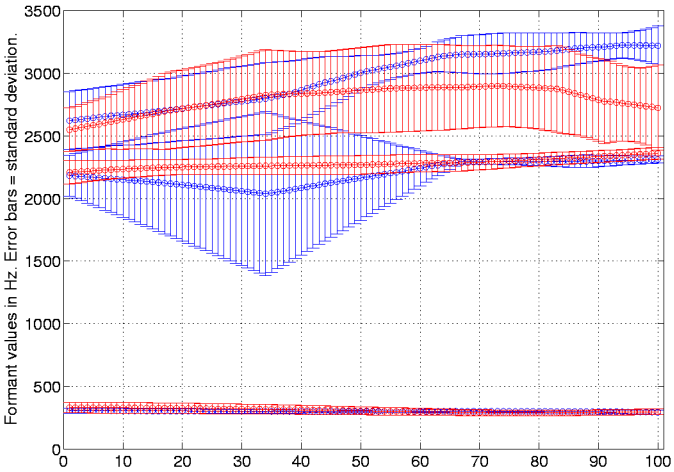


Fig. 3.19a. Condition de lecture S : en bleu, I : en rouge.

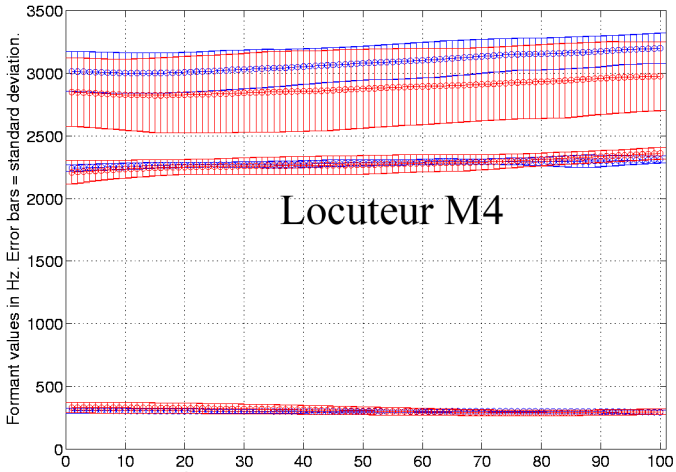


Fig. 3.19b. Condition de lecture S : en bleu, I : en rouge.

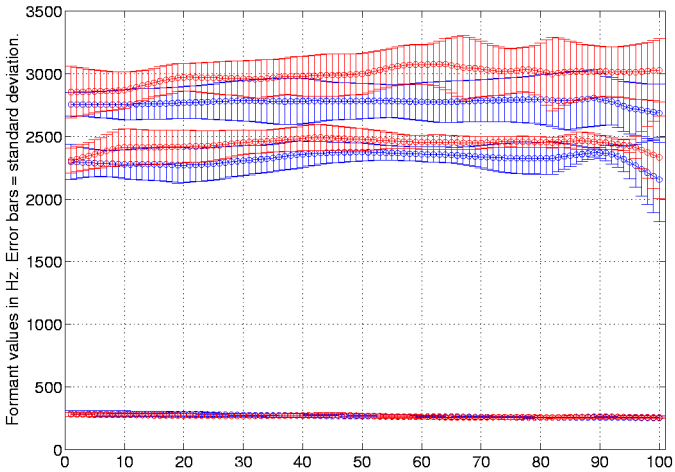


Fig. 3.20a. Condition de lecture S : en bleu, I : en rouge.

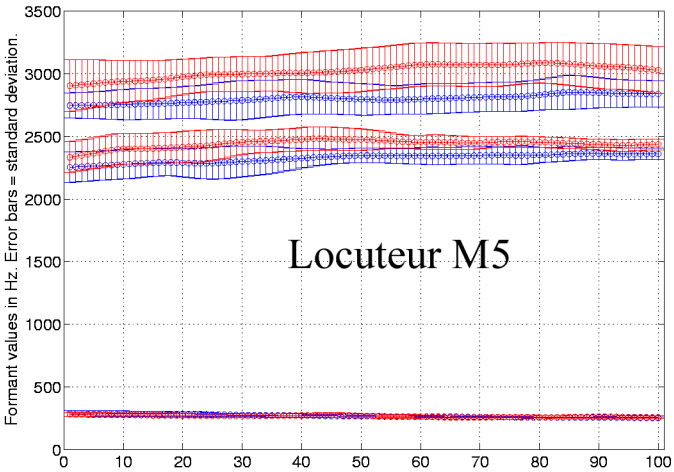


Fig. 3.20b. Condition de lecture S : en bleu, I : en rouge.

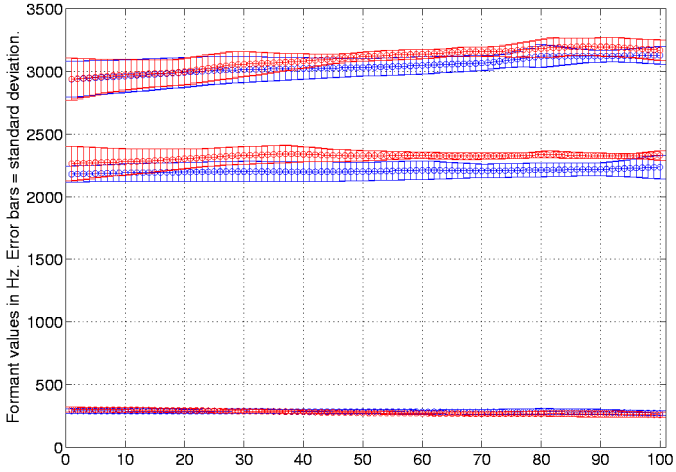


Fig. 3.21a. Condition de lecture S : en bleu, I : en rouge.

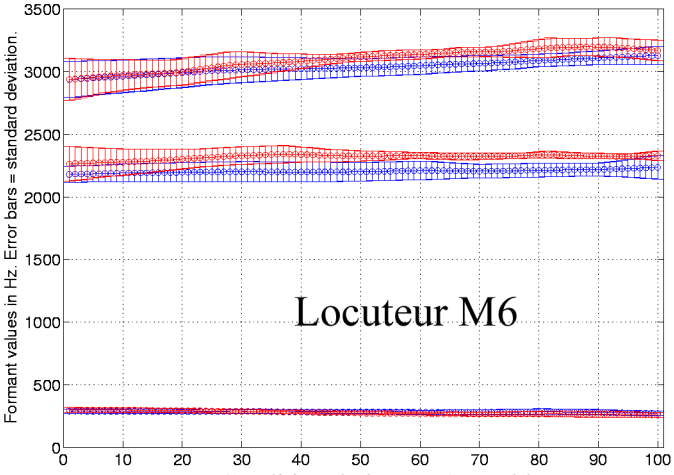


Fig. 3.21b. Condition de lecture S : en bleu, I : en rouge.

Fig. 3.13 à 3.27 : estimation des trois premiers formants des voyelles des monosyllabes en anglais.
Résultats bruts (figures 3.22a, 3.23a...) et après suppression de valeurs aberrantes suite à vérification visuelle (figures 3.22b, 3.23b...), locuteurs M3 à M7.

Colonne de gauche :
résultats bruts du logiciel Praat

Colonne de droite :
résultats vérifiés visuellement

Voyelle : / i: /

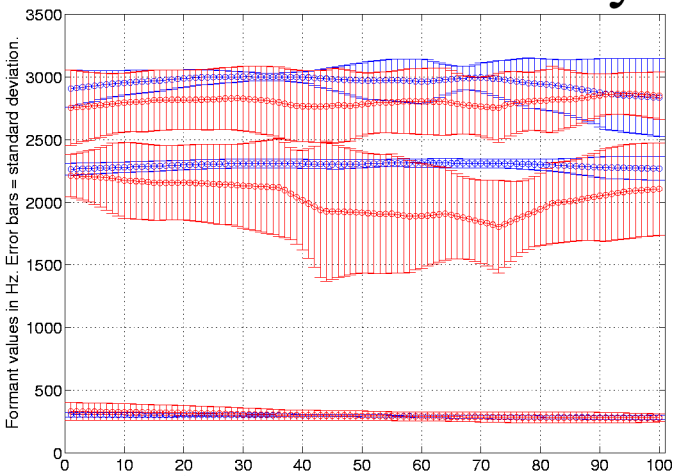


Fig. 3.22a. Condition de lecture S : en bleu, I : en rouge.

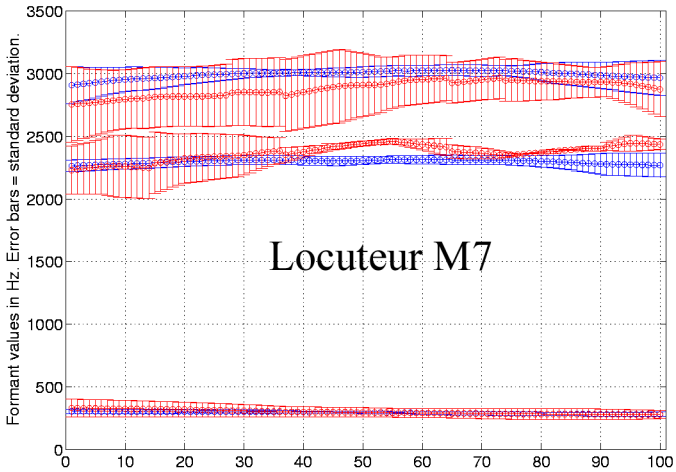


Fig. 3.22b. Condition de lecture S : en bleu, I : en rouge.

Voyelle : / u: /

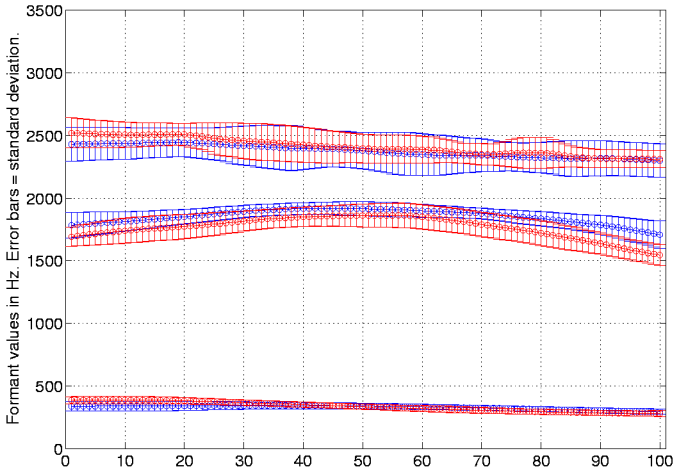


Fig. 3.23a. Condition de lecture S : en bleu, I : en rouge.

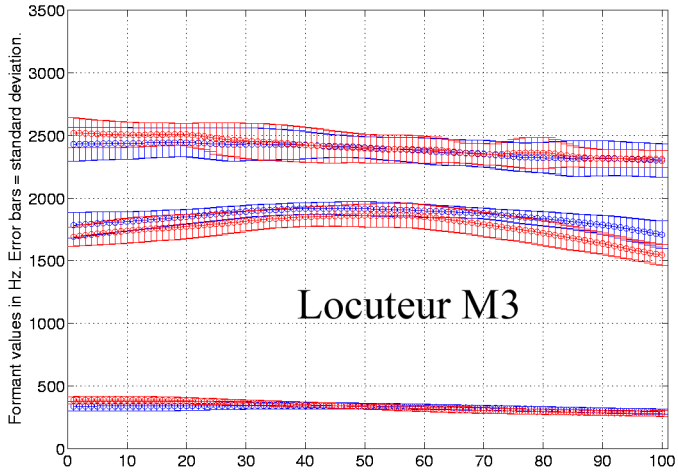


Fig. 3.23b. Condition de lecture S : en bleu, I : en rouge.

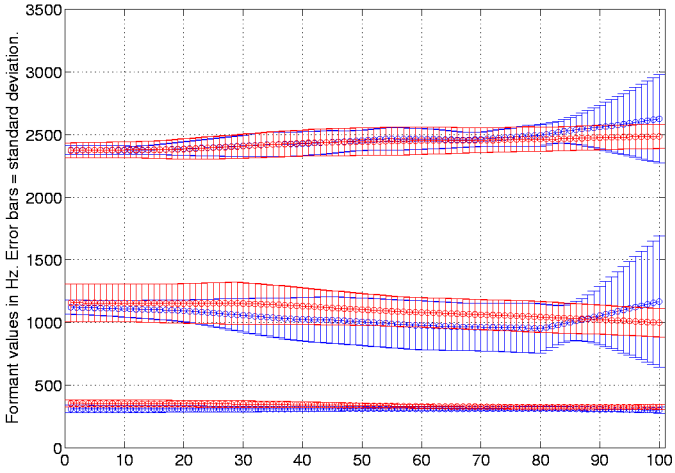


Fig. 3.24a. Condition de lecture S : en bleu, I : en rouge.

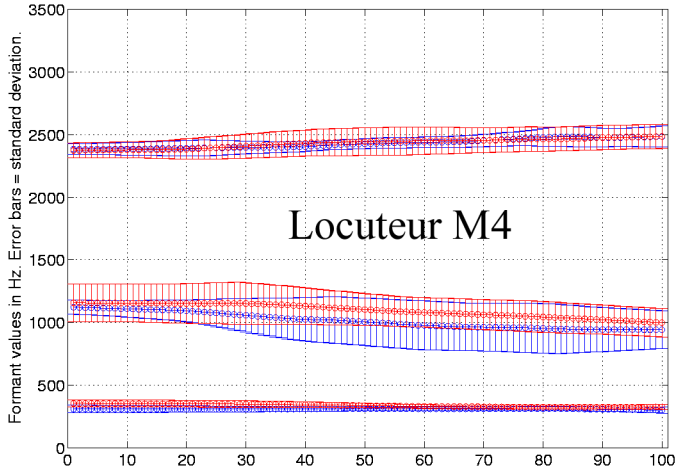


Fig. 3.24b. Condition de lecture S : en bleu, I : en rouge.

Fig. 3.13 à 3.27 : estimation des trois premiers formants des voyelles des monosyllabes en anglais.
Résultats bruts (figures 3.25a, 3.26a...) et après suppression de valeurs aberrantes suite à vérification visuelle (figures 3.25b, 3.26b...), locuteurs M3 à M7.

Voyelle : / u: /

Colonne de gauche : résultats bruts du logiciel Praat

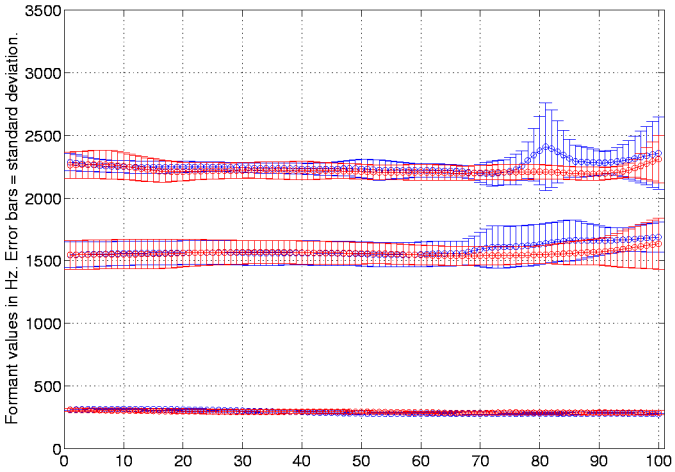


Fig. 3.25a. Condition de lecture S : en bleu, I : en rouge.

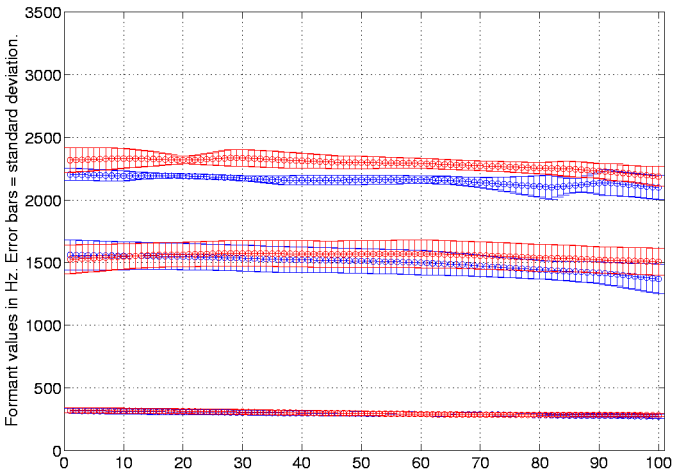


Fig. 3.26a. Condition de lecture S : en bleu, I : en rouge.

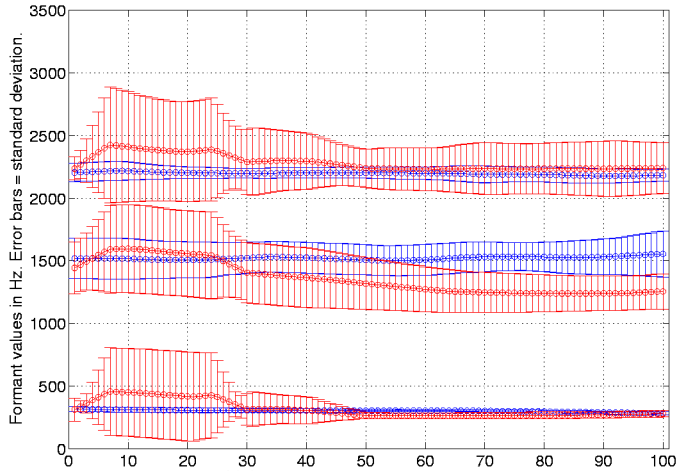


Fig. 3.27a. Condition de lecture S : en bleu, I : en rouge.

Colonne de droite : résultats vérifiés visuellement

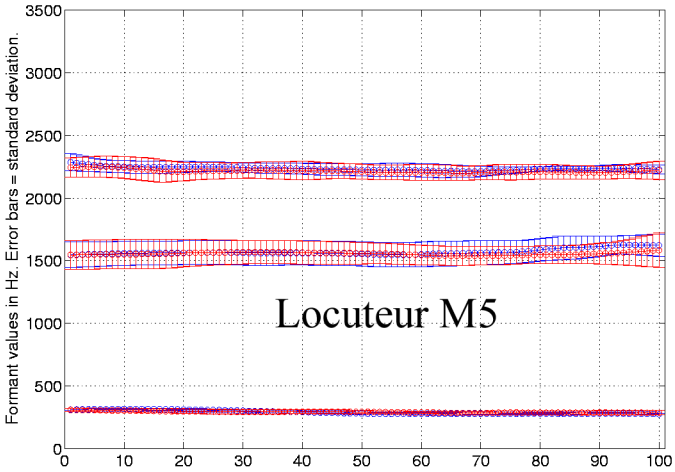


Fig. 3.25b. Condition de lecture S : en bleu, I : en rouge.

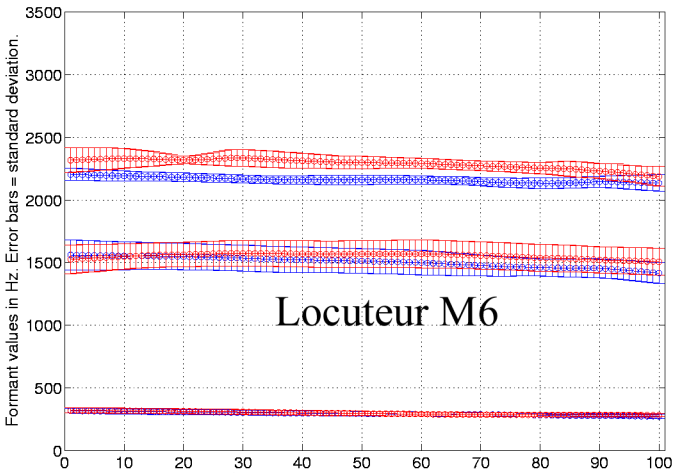


Fig. 3.26b. Condition de lecture S : en bleu, I : en rouge.

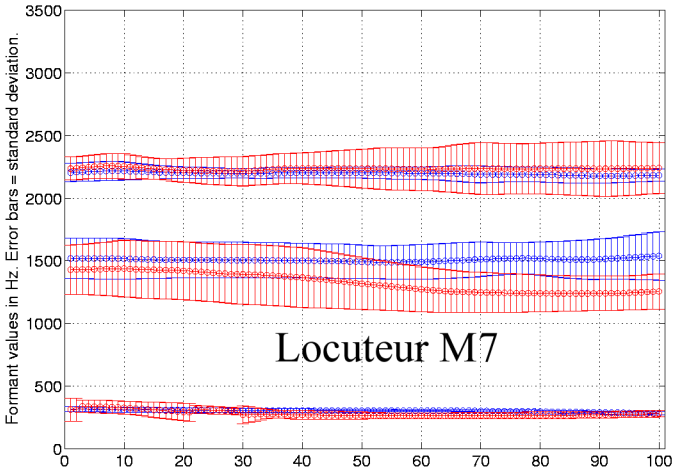


Fig. 3.27b. Condition de lecture S : en bleu, I : en rouge.

Fig. 3.28 à 3.30. Courbes moyennes d'intensité acoustique globale des mots-cible en anglais, permettant une comparaison entre conditions de lecture. Monosyllabes. Locuteur M4.

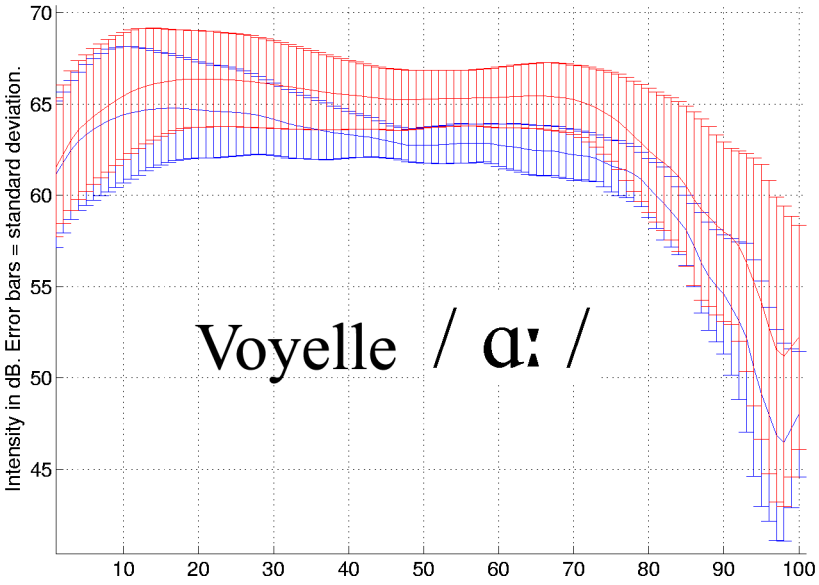


Fig. 3.28.

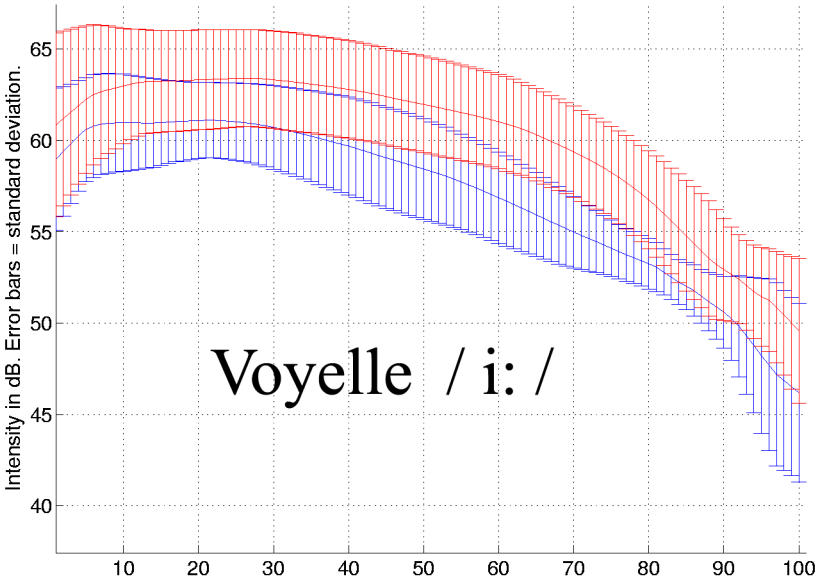


Fig. 3.29.

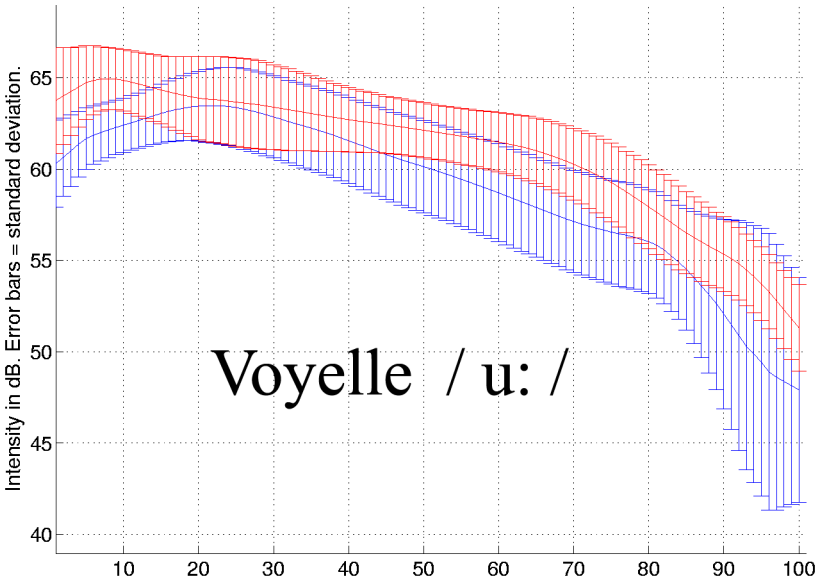


Fig. 3.30.

Figures 3.31 à 3.45 : courbes moyennes de F0 et quotient ouvert des mots-cible en naxi, permettant une comparaison entre conditions de lecture. Locuteurs M5, M7, M8, M9, et locutrice F2.

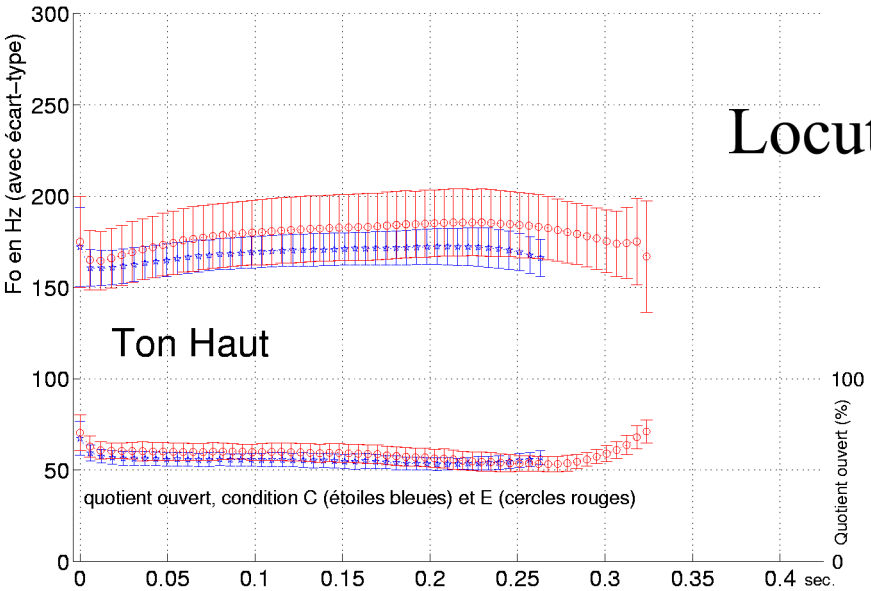


Fig. 3.31. Courbes moyennes du ton Haut, loc. M5, cond. S (bleu) et I (rouge), 84 syll.

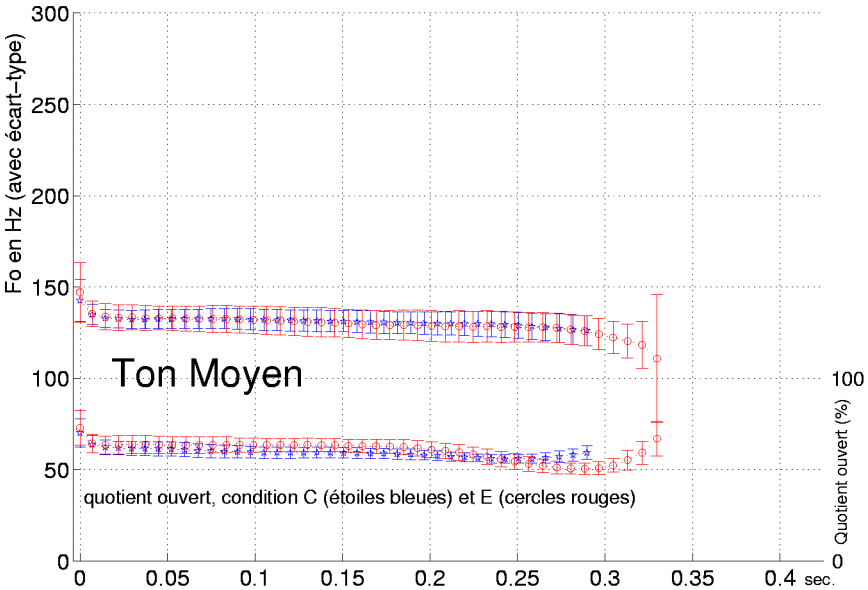


Fig. 3.32. Courbes moyennes du ton Moyen, loc. M5, cond. S (bleu) et I (rouge), 84 syll.

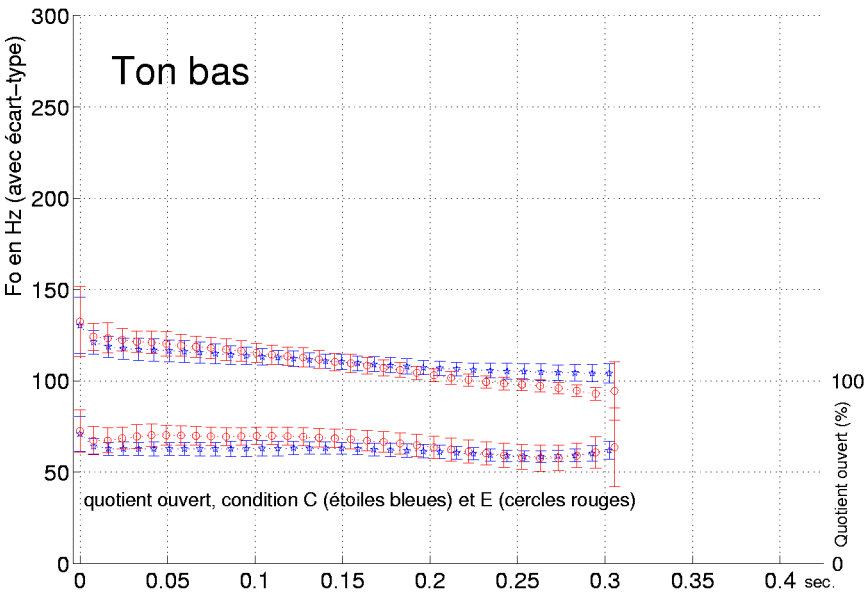


Fig. 3.33. Courbes moyennes du ton Bas, loc. M5, cond. S (bleu) et I (rouge), 84 syll.

Figures 3.31 à 3.45 : courbes moyennes de F0 et quotient ouvert des mots-cible en naxi, permettant une comparaison entre conditions de lecture. Locuteurs M5, M7, M8, M9, et locutrice F2.

Locuteur M7

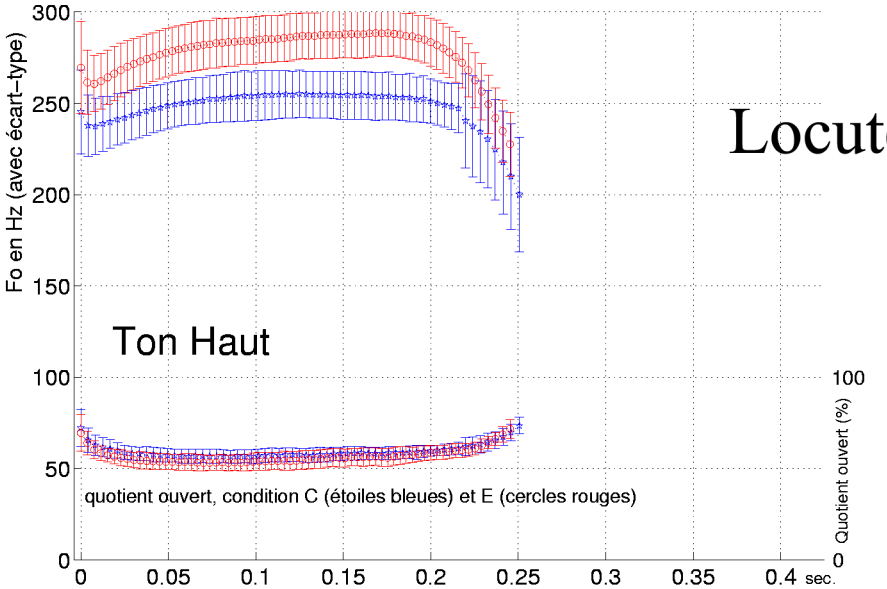


Fig. 3.34. Courbes moyennes du ton Haut, loc. M7, cond. S (bleu) et I (rouge), 84 syll.

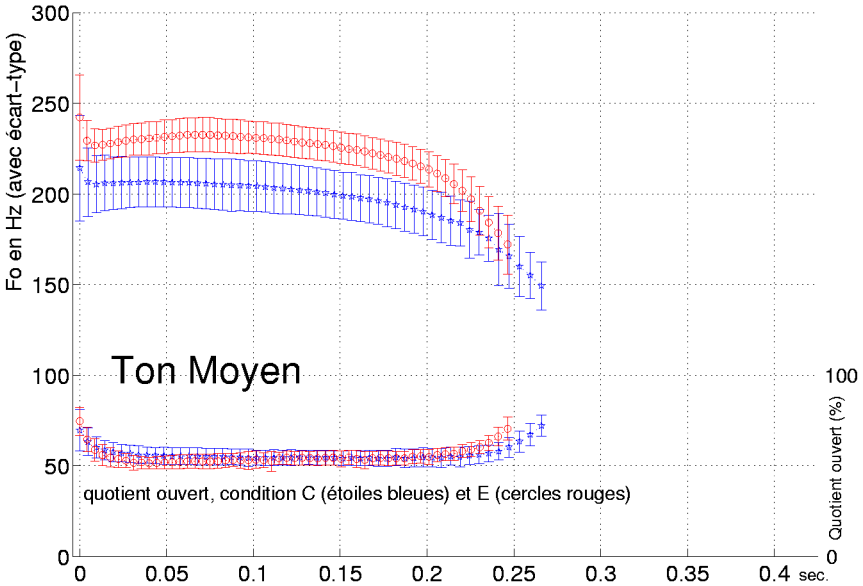


Fig. 3.35. Courbes moyennes du ton Moyen, loc. M7, cond. S (bleu) et I (rouge), 84 syll.

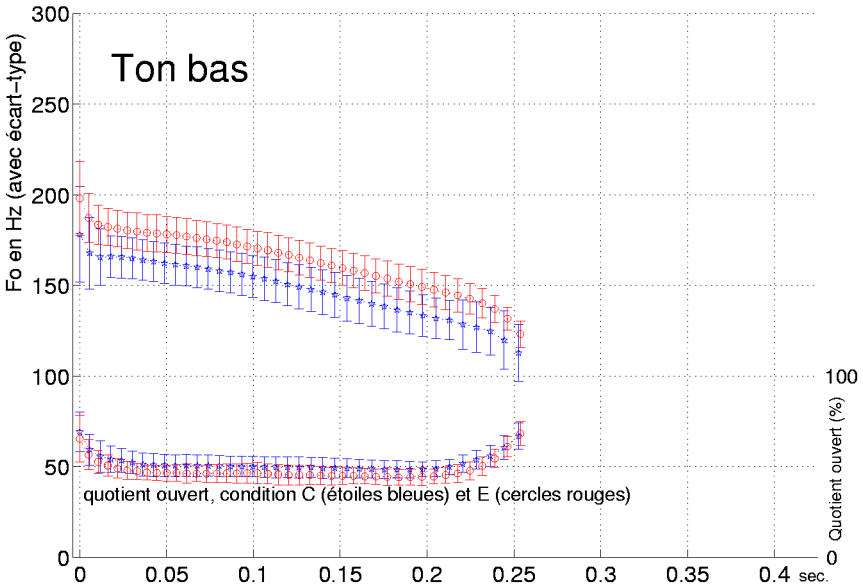


Fig. 3.36. Courbes moyennes du ton Bas, loc. M7, cond. S (bleu) et I (rouge), 84 syll.

Figures 3.31 à 3.45 : courbes moyennes de F0 et quotient ouvert des mots-cible en naxi, permettant une comparaison entre conditions de lecture. Locuteurs M5, M7, M8, M9, et locutrice F2.

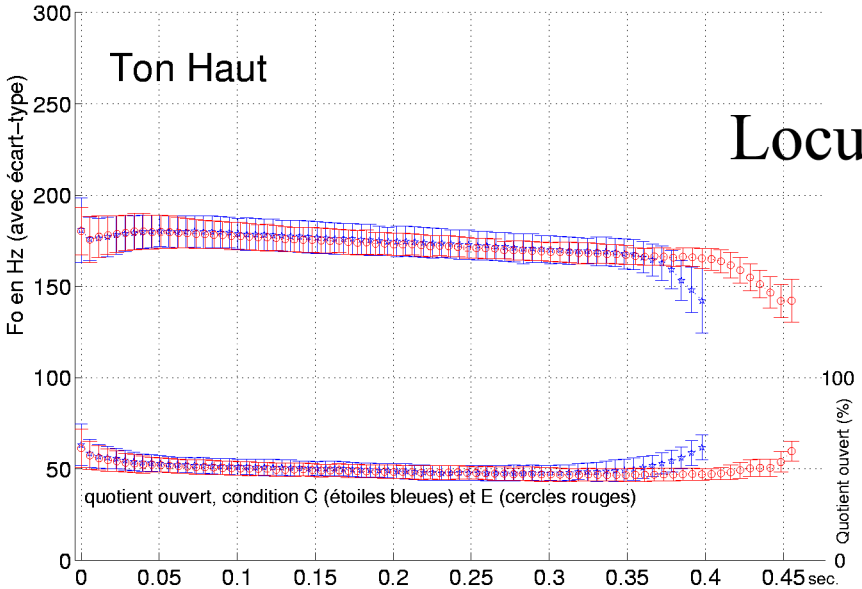


Fig. 3.37. Courbes moyennes du ton Haut, loc. M8, cond. S (bleu) et I (rouge), 84 syll.

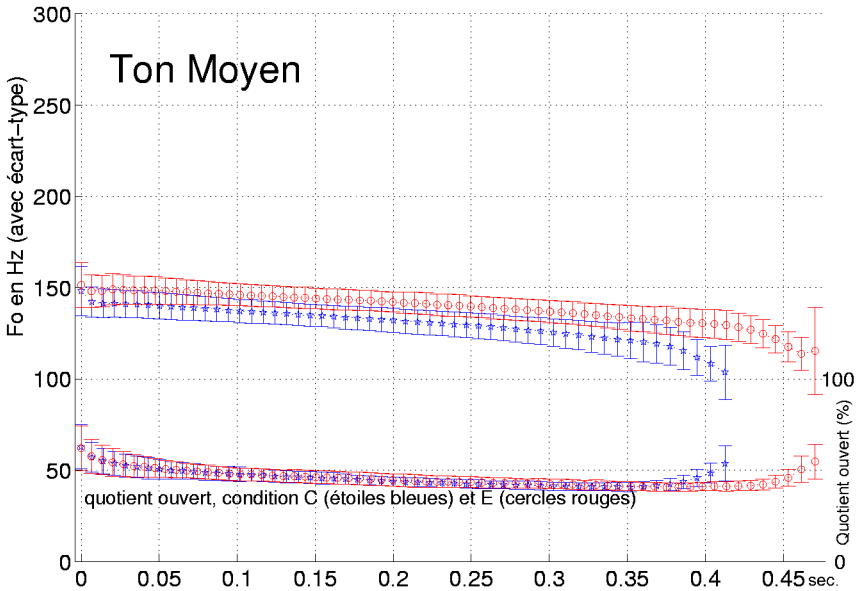


Fig. 3.38. Courbes moyennes du ton Moyen, loc. M8, cond. S (bleu) et I (rouge), 84 syll.

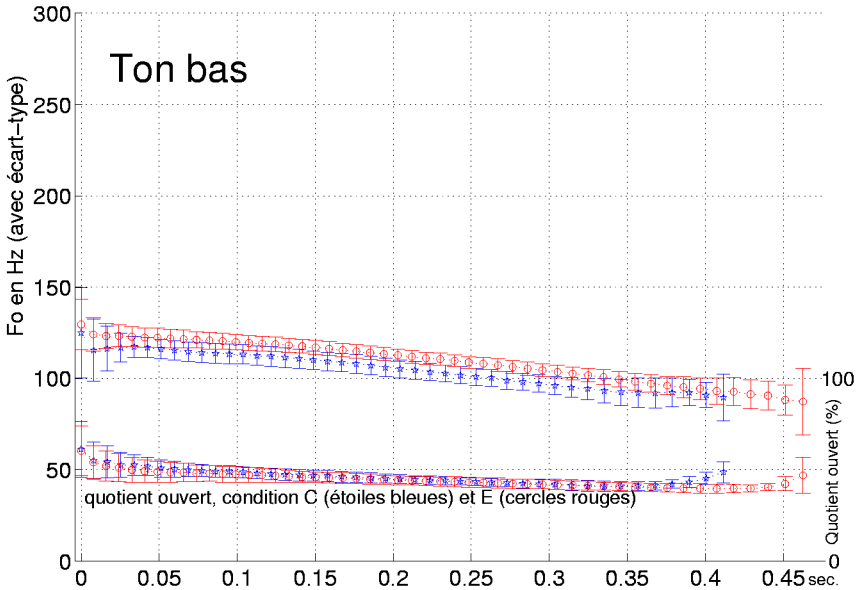


Fig. 3.39. Courbes moyennes du ton Bas, loc. M8, cond. S (bleu) et I (rouge), 84 syll.

Figures 3.31 à 3.45 : courbes moyennes de F0 et quotient ouvert des mots-cible en naxi, permettant une comparaison entre conditions de lecture. Locuteurs M5, M7, M8, M9, et locutrice F2.

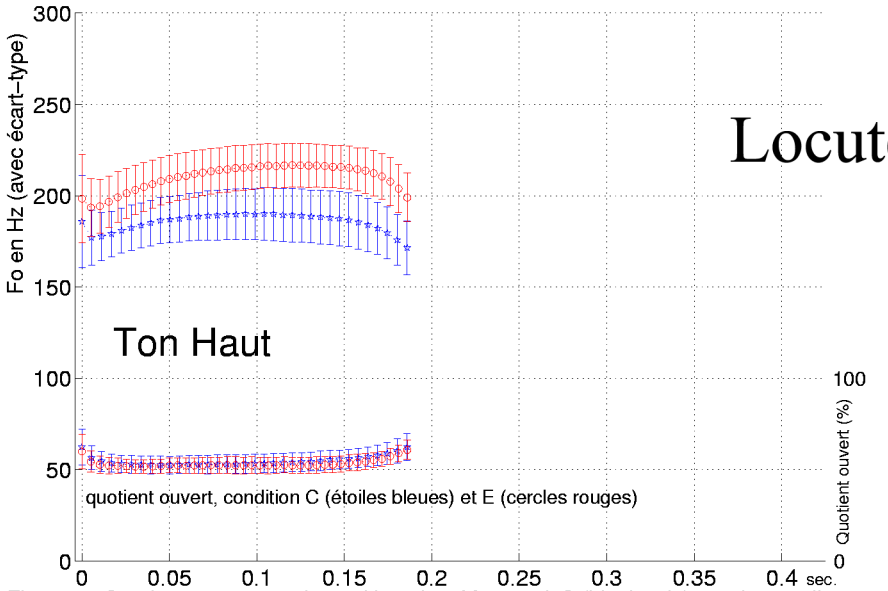


Fig. 3.40. Courbes moyennes du ton Haut, loc. M9, cond. S (bleu) et I (rouge), 84 syll.

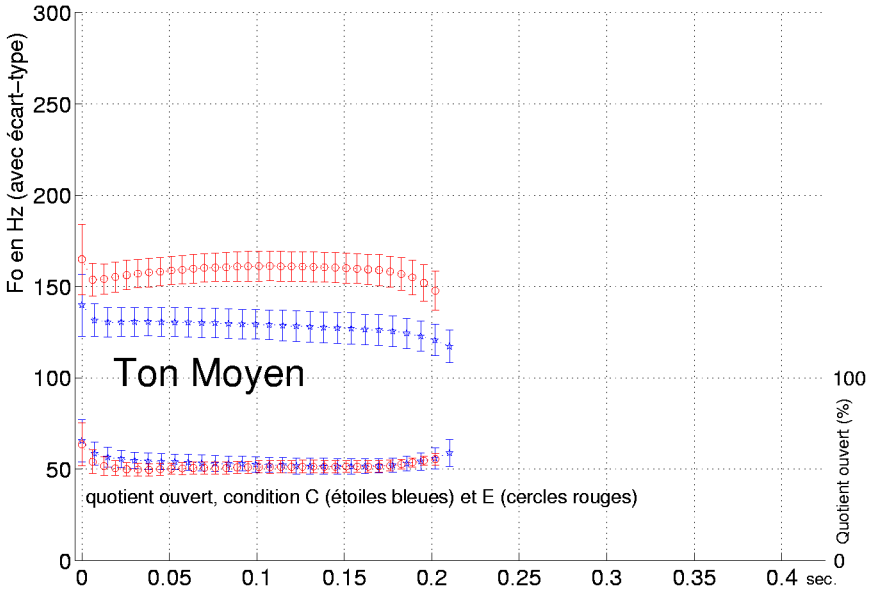


Fig. 3.41. Courbes moyennes du ton Moyen, loc. M9, cond. S (bleu) et I (rouge), 84 syll.

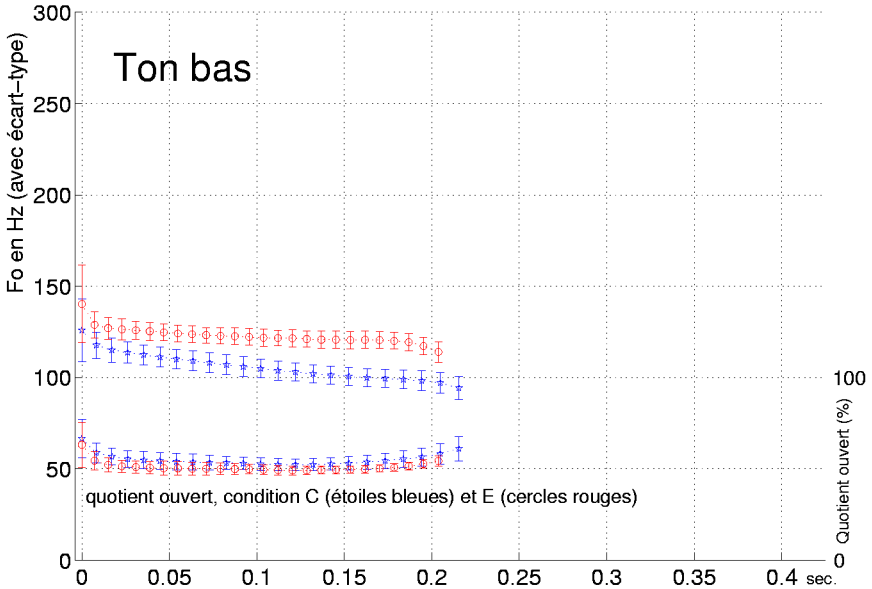


Fig. 3.42. Courbes moyennes du ton Bas, loc. M9, cond. S (bleu) et I (rouge), 84 syll.

Locutrice F2

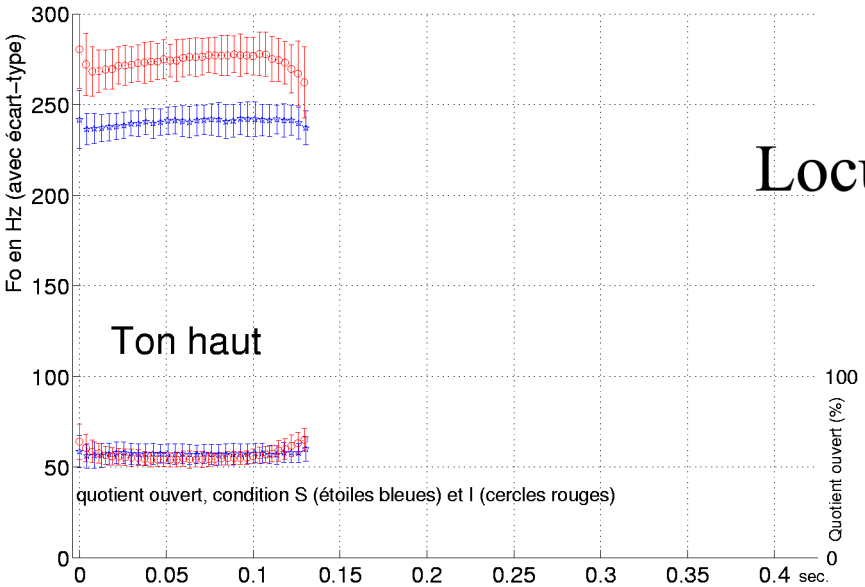


Fig. 3.43. Courbes moyennes du ton Haut, locutrice F2, conditions S et I .

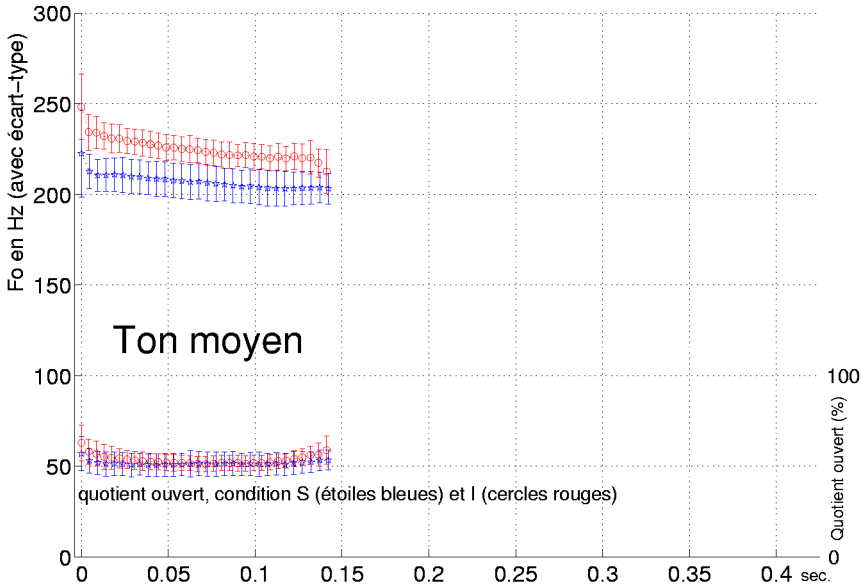


Fig. 3.44. Courbes moyennes du ton Moyen, locutrice F2, conditions S et I.

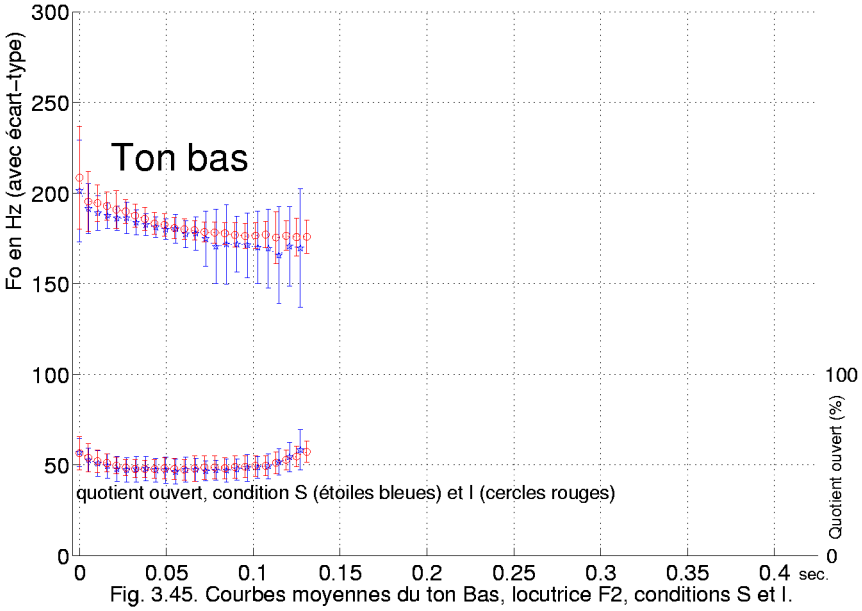


Fig. 3.45. Courbes moyennes du ton Bas, locutrice F2, conditions S et I.

Figures 3.31 à 3.45 : courbes moyennes de F0 et quotient ouvert des mots-cible en naxi, permettant une comparaison entre conditions de lecture. Locuteurs M5, M7, M8, M9, et locutrice F2.

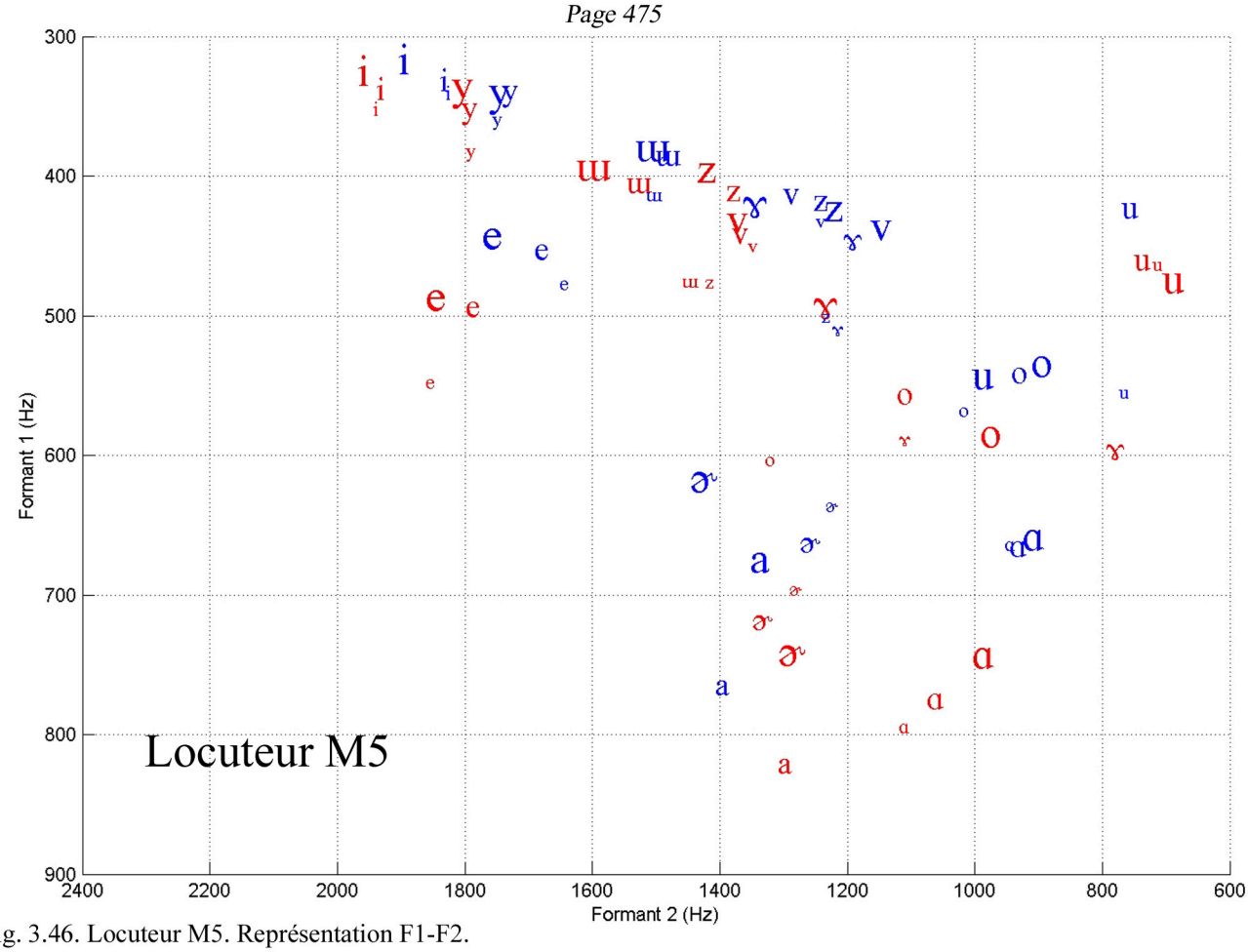


Fig. 3.46. Locuteur M5. Représentation F1-F2.

Fig. 3.46 à 3.55 : formants des voyelles naxi (dans un espace F1-F2 sur les figures à numéro impair, F2-F3 sur les figures à numéro pair), contrastant la condition de lecture soignée (en bleu) et la condition de lecture insistante (en rouge) ainsi que les trois tons : voyelles au ton H représentées par un caractère de grande taille, au ton M par un caractère de taille moyenne, au ton L par un caractère de petite taille.

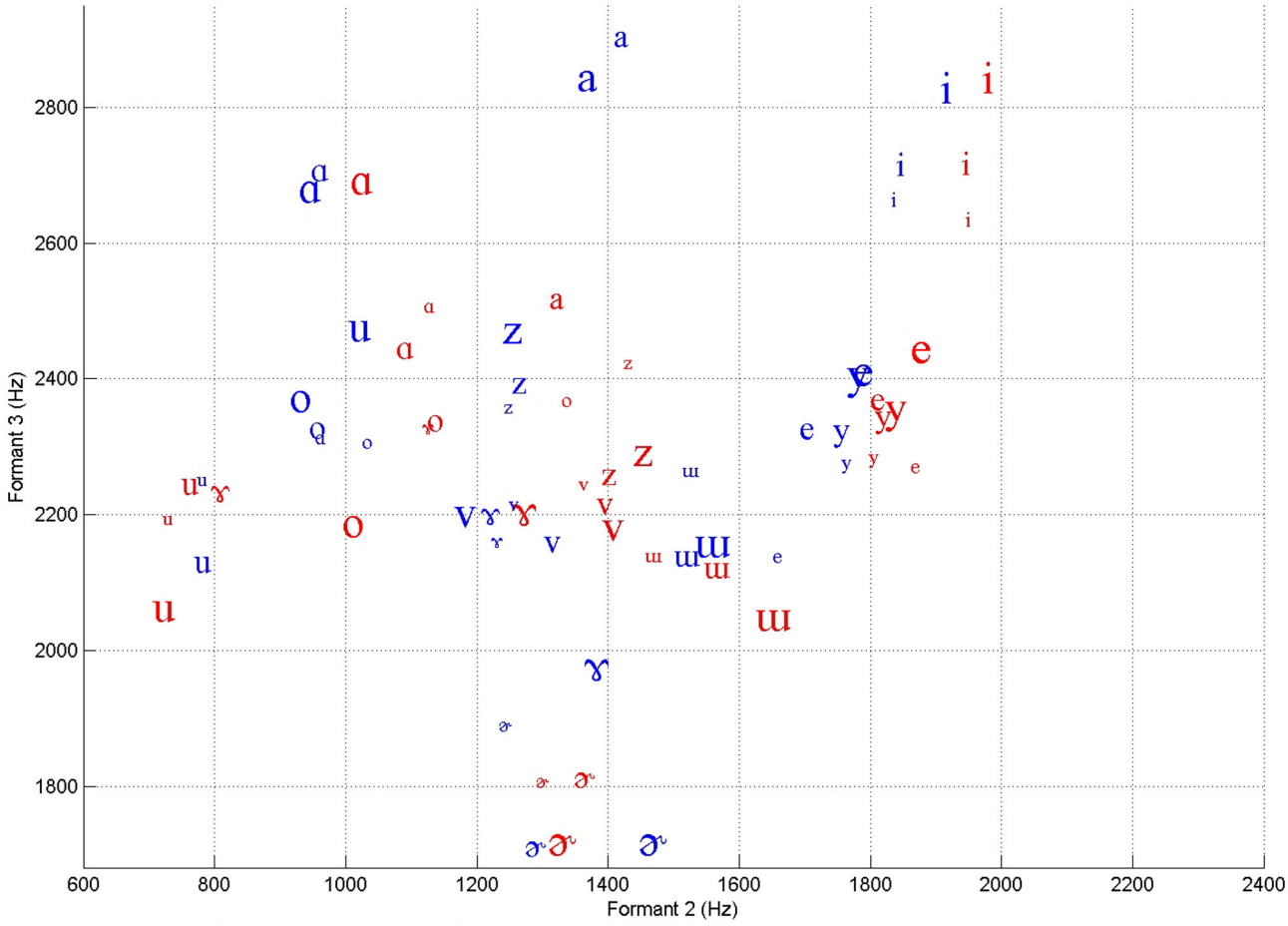


Fig. 3.47. Locuteur M5. Représentation F2-F3.

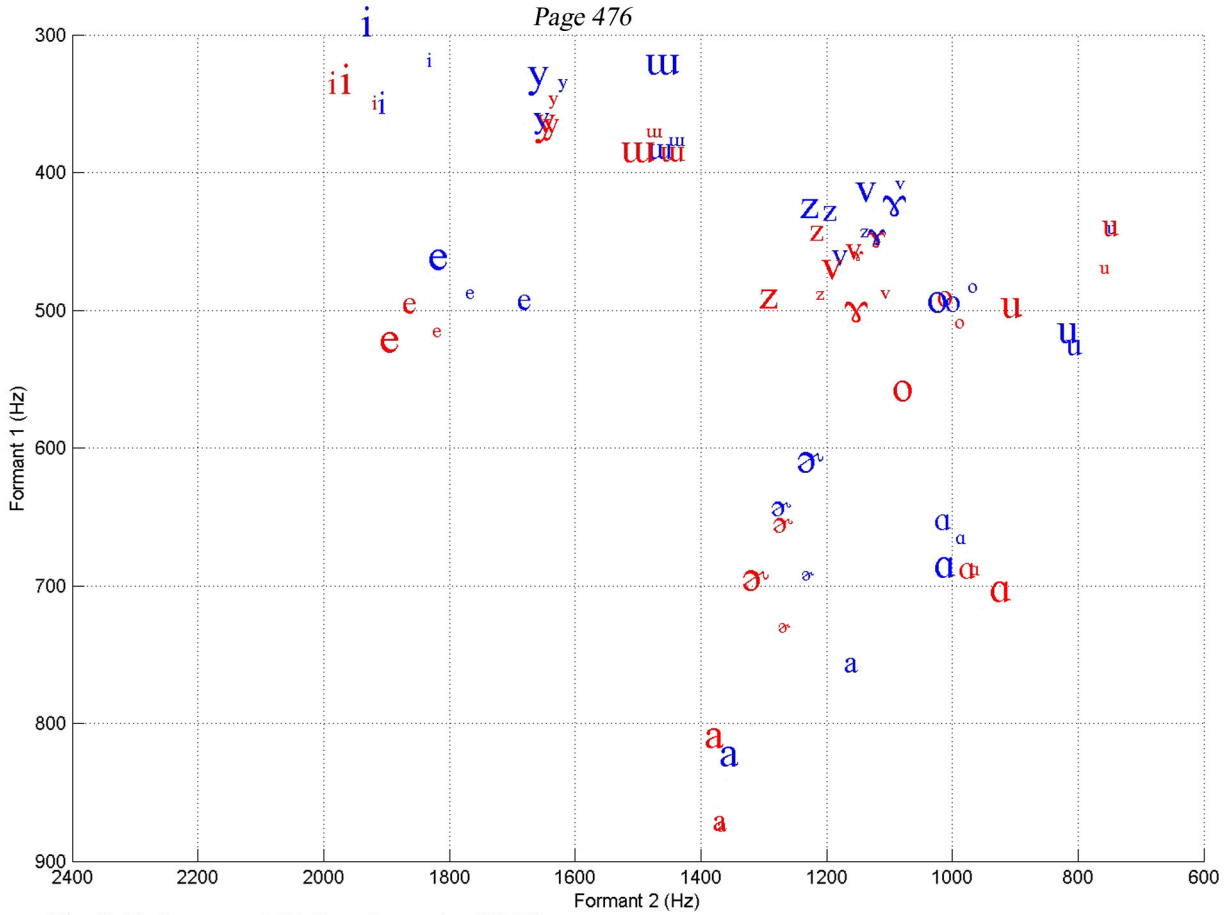


Fig. 3.48. Locuteur M7. Représentation F1-F2.

Fig. 3.46 à 3.55 : formants des voyelles naxi (dans un espace F1-F2 sur les figures à numéro impair, F2-F3 sur les figures à numéro pair), contrastant la condition de lecture soignée (en bleu) et la condition de lecture insistante (en rouge) ainsi que les trois tons : voyelles au ton H représentées par un caractère de grande taille, au ton M par un caractère de taille moyenne, au ton L par un caractère de petite taille.

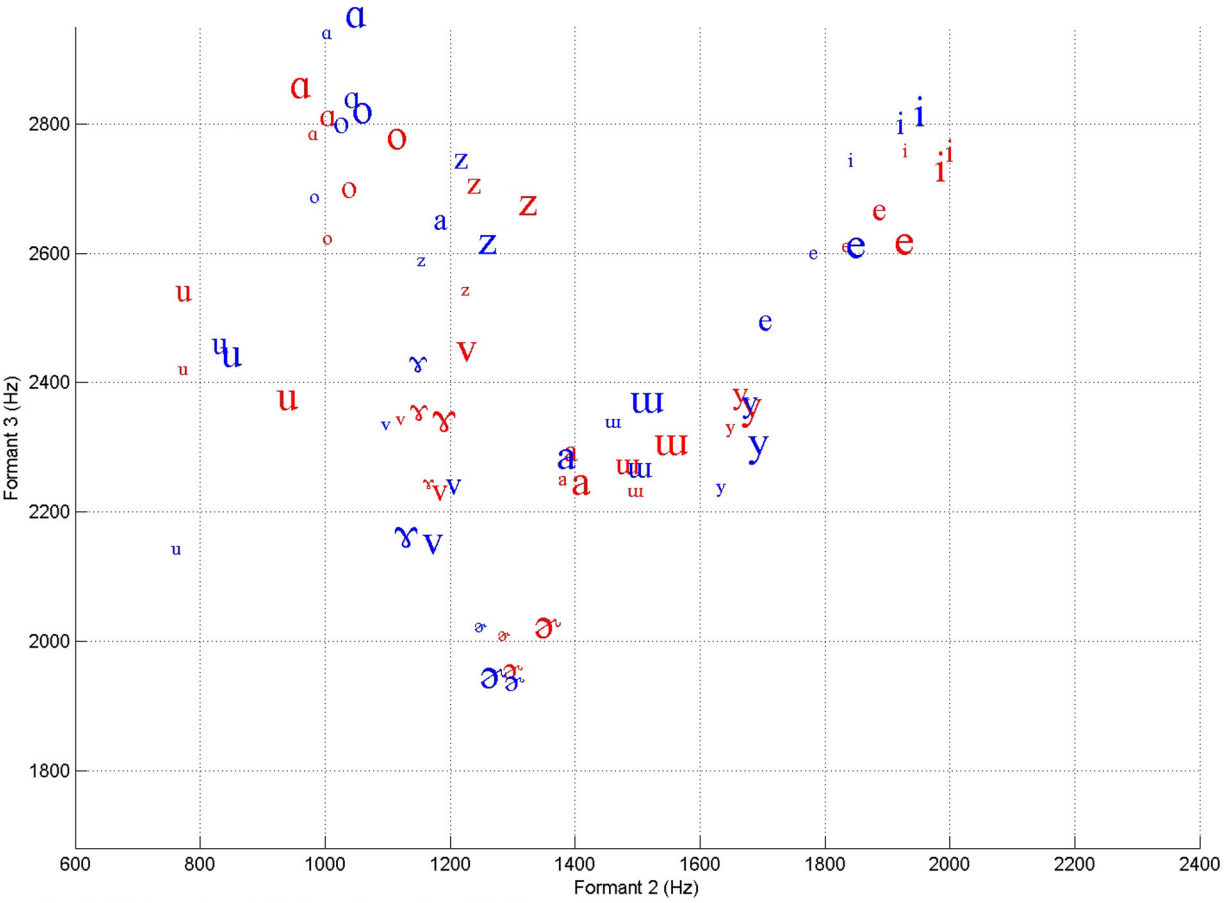


Fig. 3.49. Locuteur M7. Représentation F2-F3.

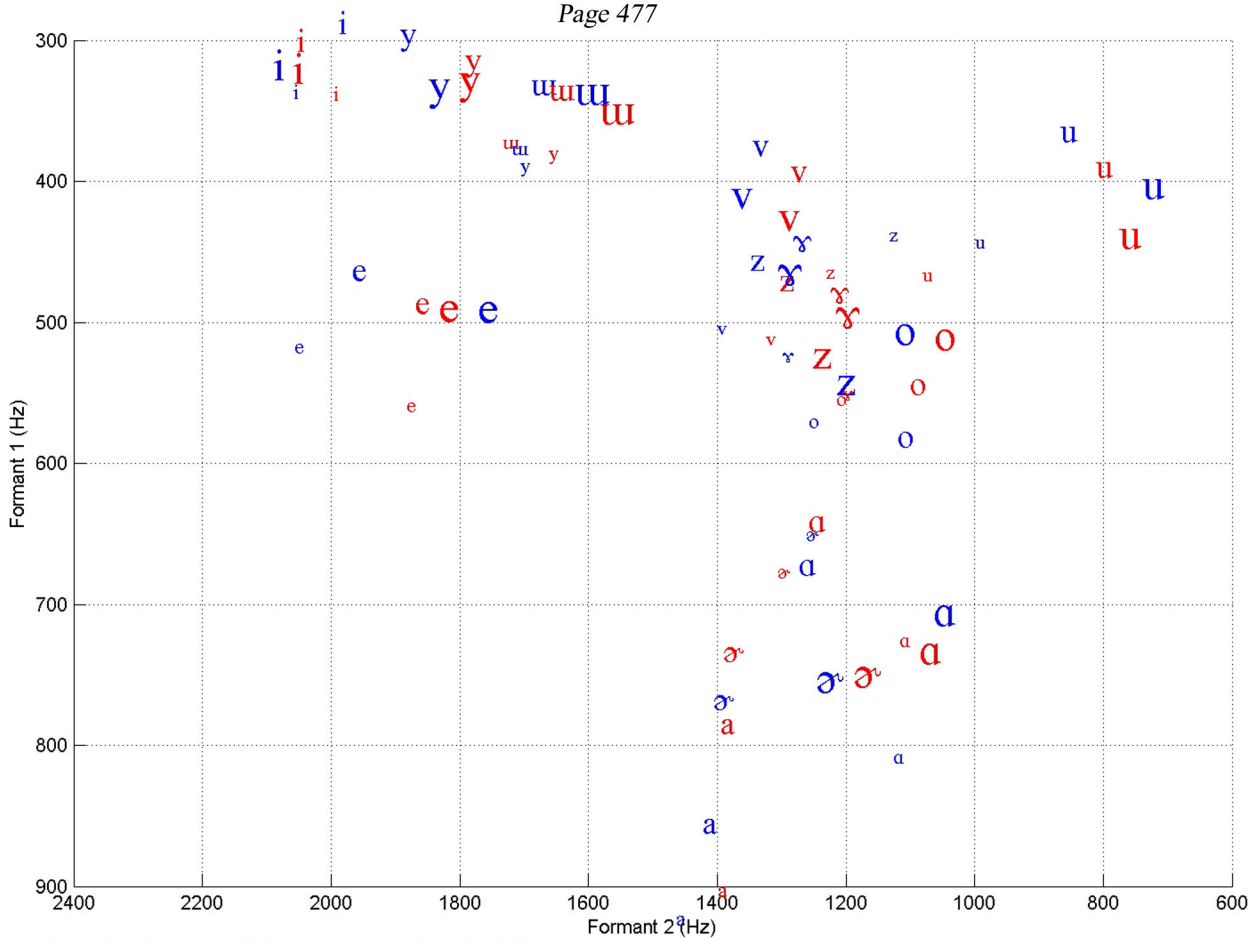


Fig. 3.50. Locuteur M8. Représentation F1-F2.

Fig. 3.46 à 3.55 : formants des voyelles naxi (dans un espace F1-F2 sur les figures à numéro impair, F2-F3 sur les figures à numéro pair), contrastant la condition de lecture soignée (en bleu) et la condition de lecture insistante (en rouge) ainsi que les trois tons : voyelles au ton H représentées par un caractère de grande taille, au ton M par un caractère de taille moyenne, au ton L par un caractère de petite taille.

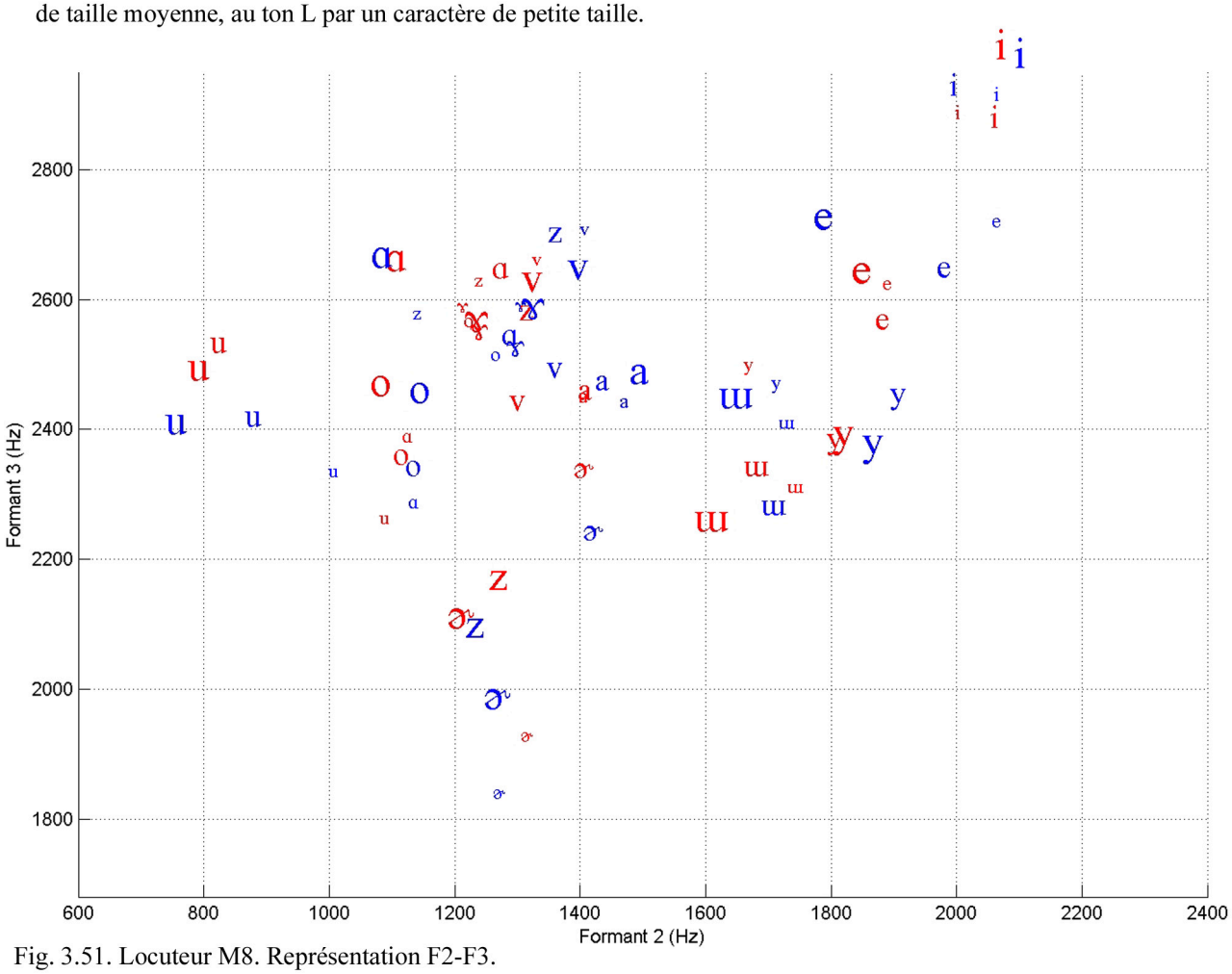


Fig. 3.51. Locuteur M8. Représentation F2-F3.

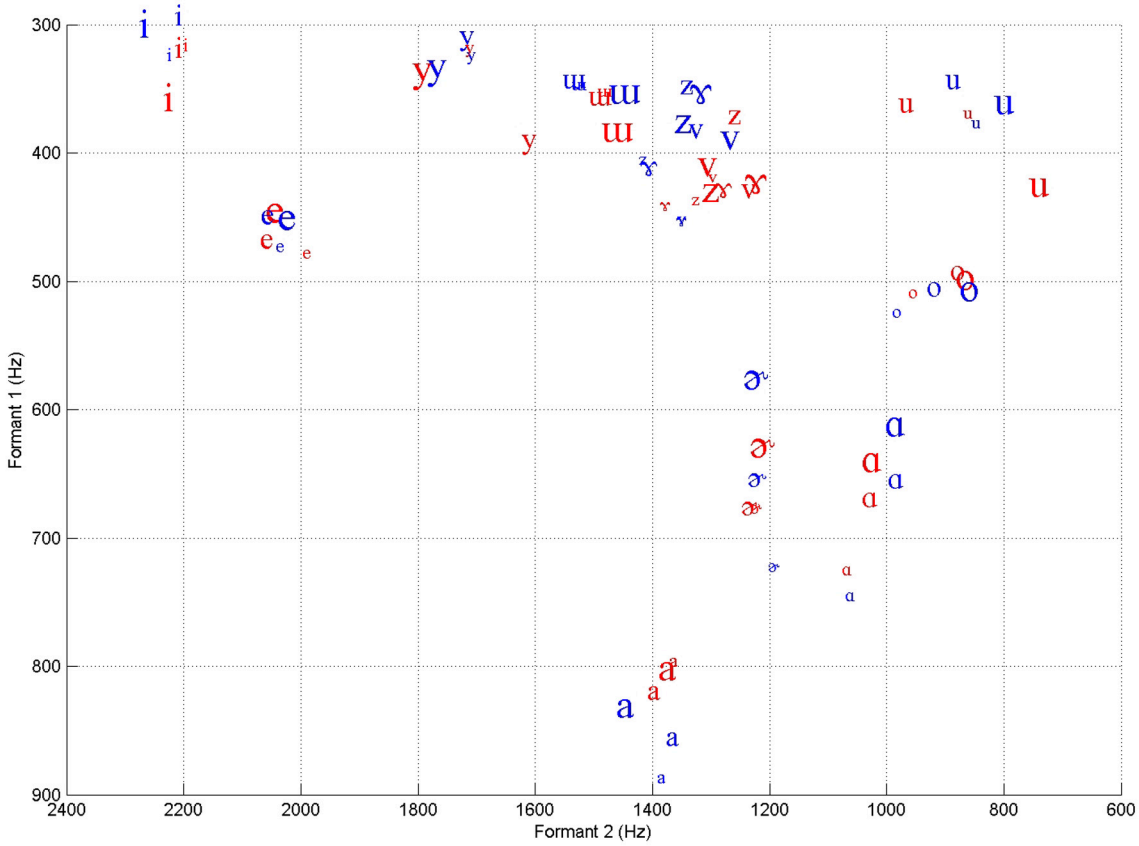


Fig. 3.52. Locuteur M9.
Représentation F1-F2.

Fig. 3.46 à 3.55 : formants des voyelles naxi (dans un espace F1-F2 sur les figures à numéro impair, F2-F3 sur les figures à numéro pair), contrastant la condition de lecture soignée (en bleu) et la condition de lecture insistante (en rouge) ainsi que les trois tons : voyelles au ton H représentées par un caractère de grande taille, au ton M par un caractère de taille moyenne, au ton L par un caractère de petite taille.

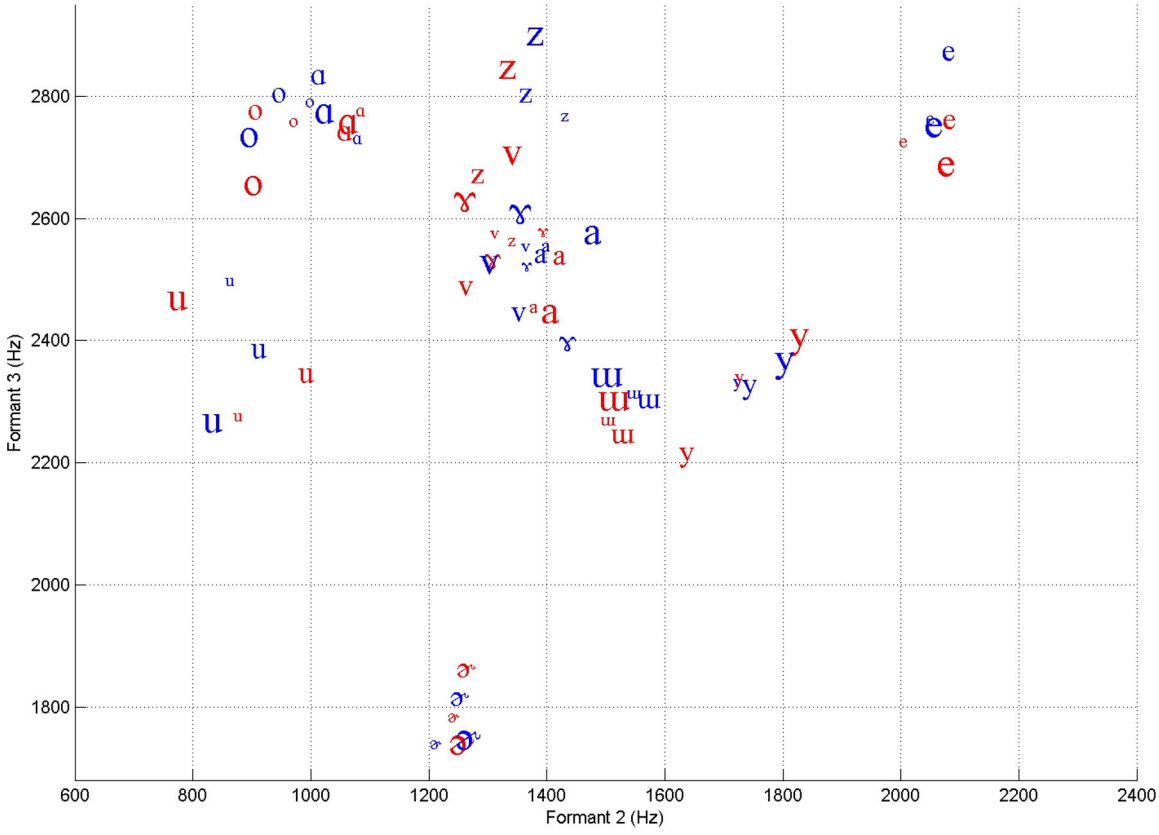


Fig. 3.53. Locuteur M9.
Représentation F2-F3.

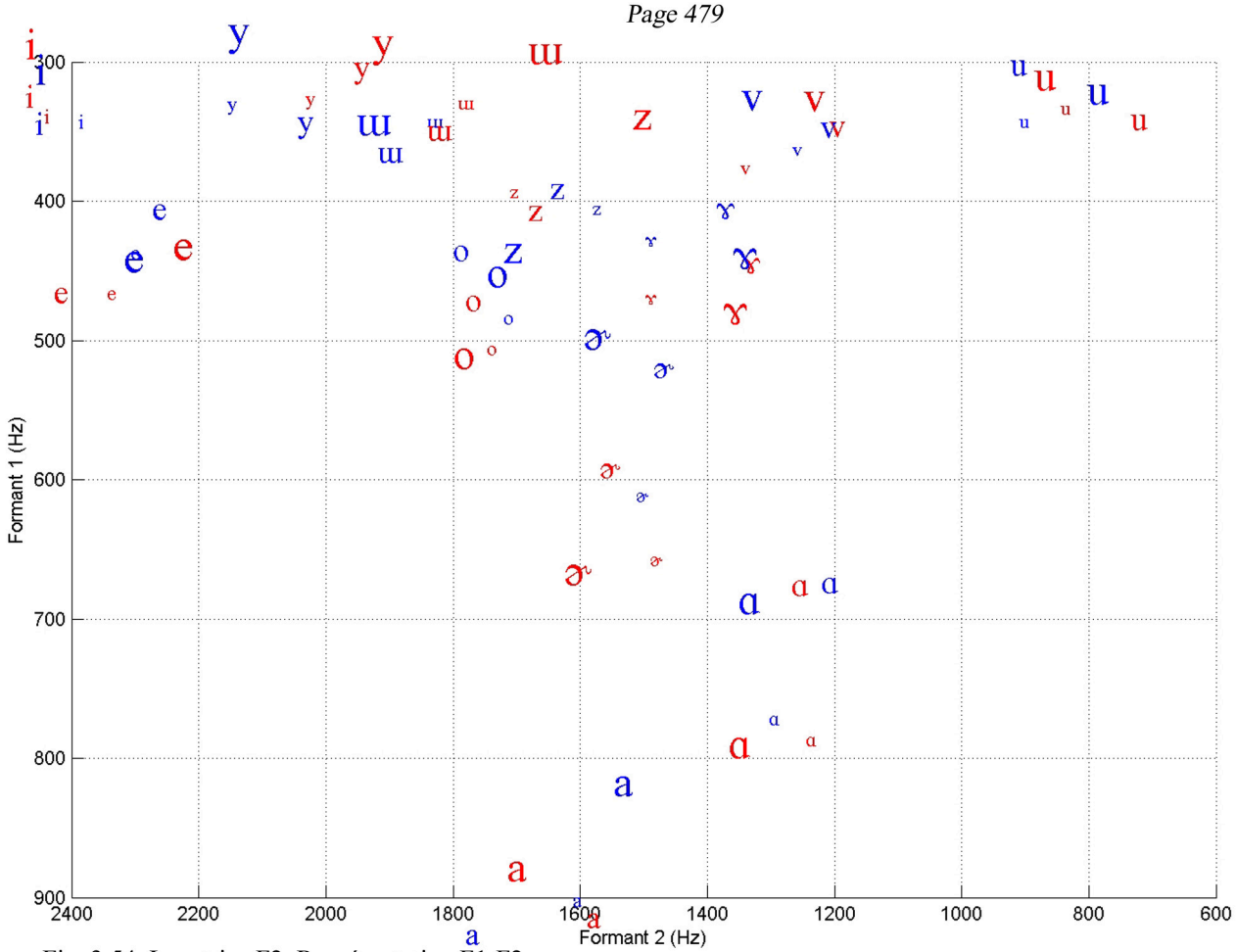


Fig. 3.54. Locutrice F2. Représentation F1-F2.

Fig. 3.46 à 3.55 : formants des voyelles naxi (dans un espace F1-F2 sur les figures à numéro impair, F2-F3 sur les figures à numéro pair), contrastant la condition de lecture soignée (en bleu) et la condition de lecture insistante (en rouge) ainsi que les trois tons : voyelles au ton H représentées par un caractère de grande taille, au ton M par un caractère de taille moyenne, au ton L par un caractère de petite taille.

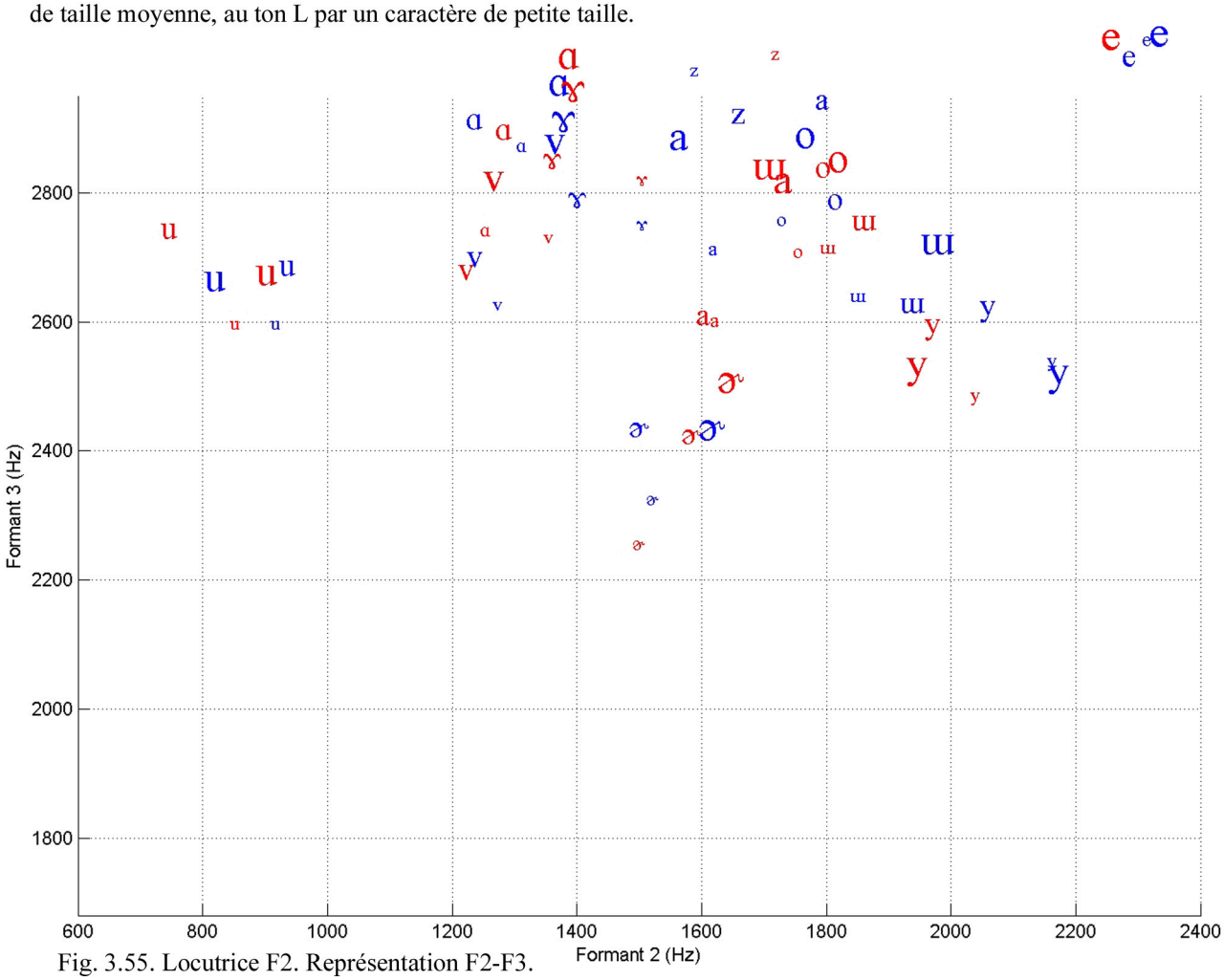


Fig. 3.55. Locutrice F2. Représentation F2-F3.

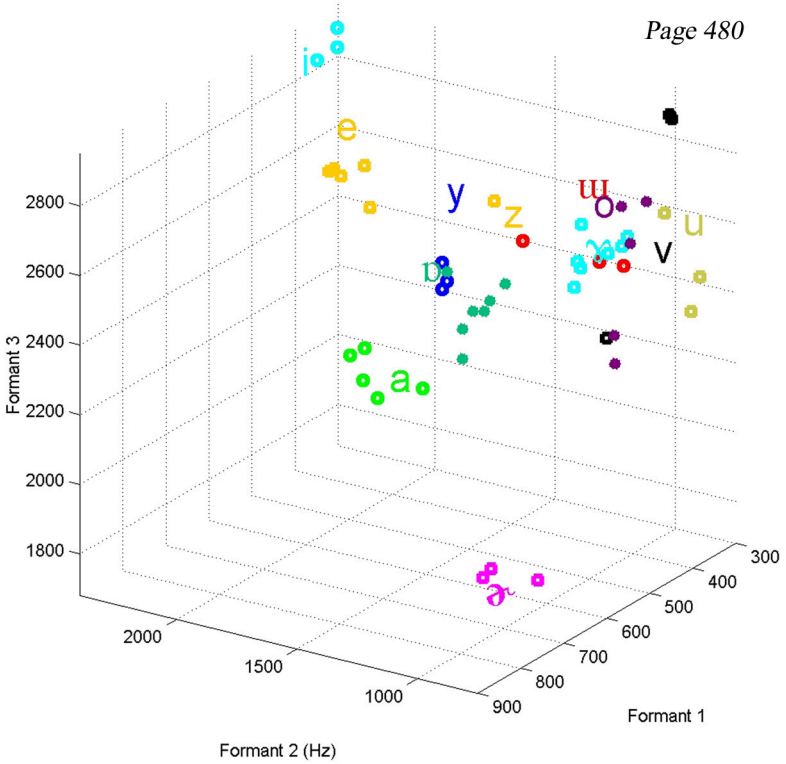


Fig. 3.56. Exemple de représentation de voyelles en trois dimensions (F1, F2, F3).
Locuteur naxi M3.

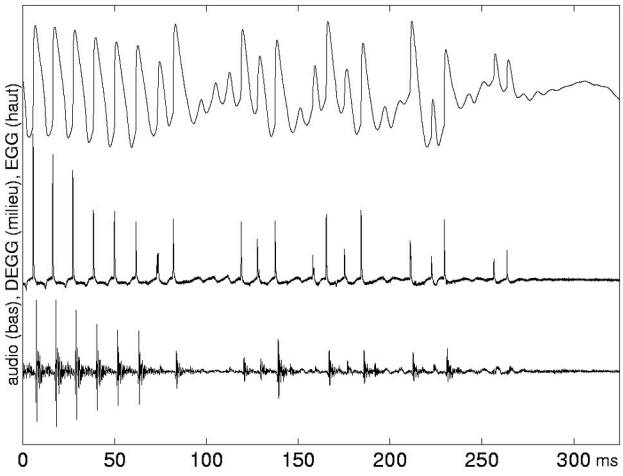


Figure 3.58. Exemple de laryngalisation (en fin de ton C1).

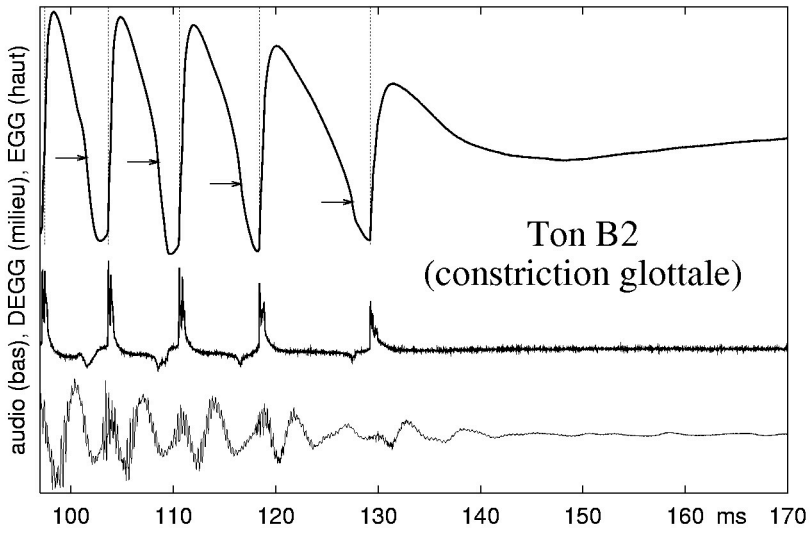


Fig. 3.57. Fin du voisement au ton B2. Lignes: fermetures, flèches: ouvertures.
EGG (haut), DEGG (milieu), audio (bas).

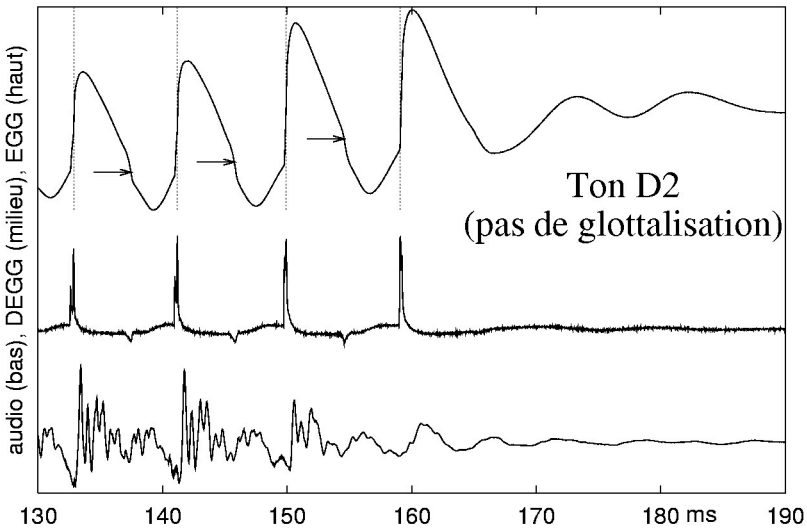


Fig. 3.59. Fin du voisement au ton D2. Lignes: fermetures, flèches: ouvertures.
EGG (haut), DEGG (milieu), audio (bas).

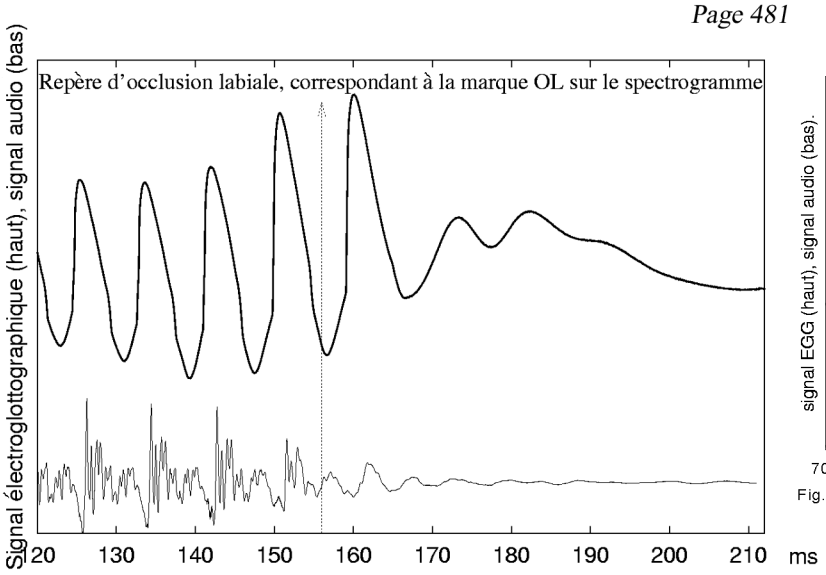


Fig. 3.60. Même exemple que sur la fig. 3.59 : fin de syllabe /ap/, locuteur M1.

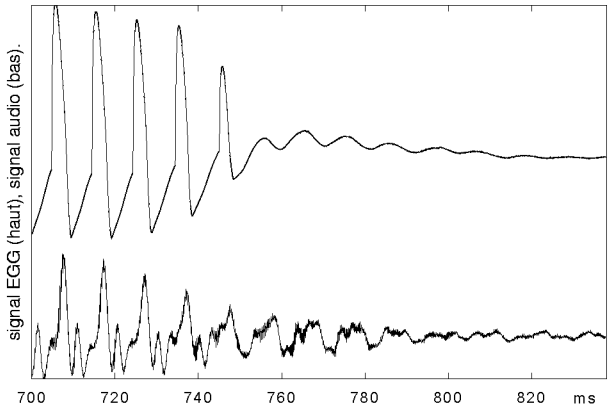
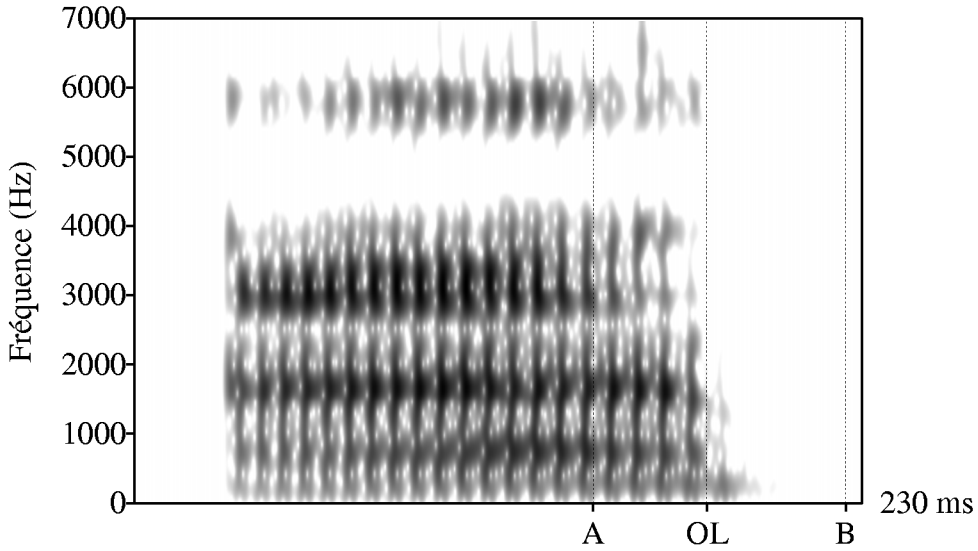
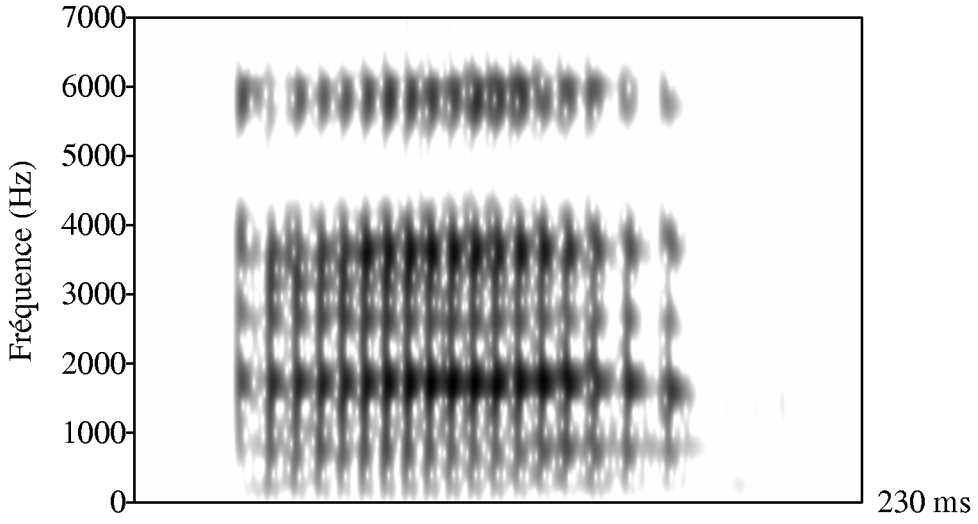


Fig. 3.61. Dernières périodes de la syllabe /i/, ton A2, avant pause. Locutrice F1. Sept petites oscillations sur signal EGG en fin de voisement.



Syllabe /ap/, ton D2. OL: repère correspondant approximativement à l'occlusion labiale. A et B: début et fin du signal électroglottographique représenté en haut de page.



Syllabe /am/, ton B2 (bas, constriction glottique). (Données du locuteur M1.)

Fig. 3.62. Spectrogrammes de syllabes aux tons B2 et D2.

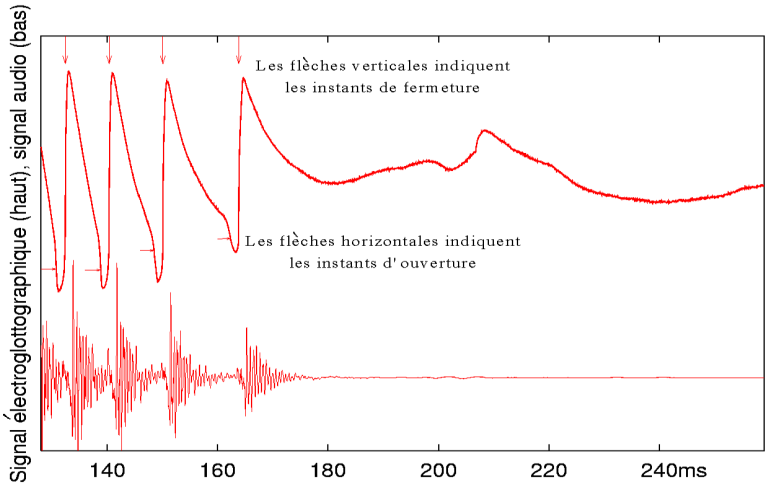


Fig. 3.63. Fin de la syllable /am/ au ton B2, locuteur M1. Condition I.
Les indications des instants de fermeture et d'ouverture proviennent du signal DEGG.

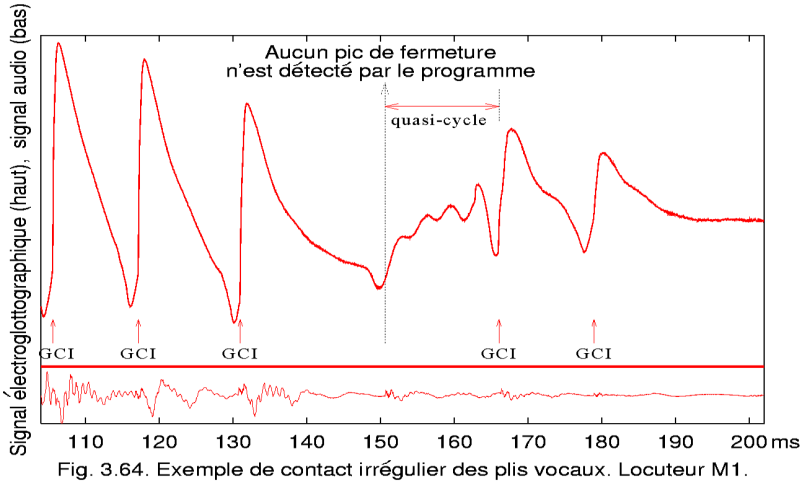


Fig. 3.64. Exemple de contact irrégulier des plis vocaux. Locuteur M1.
Syllabe avec nasale finale, ton B2.

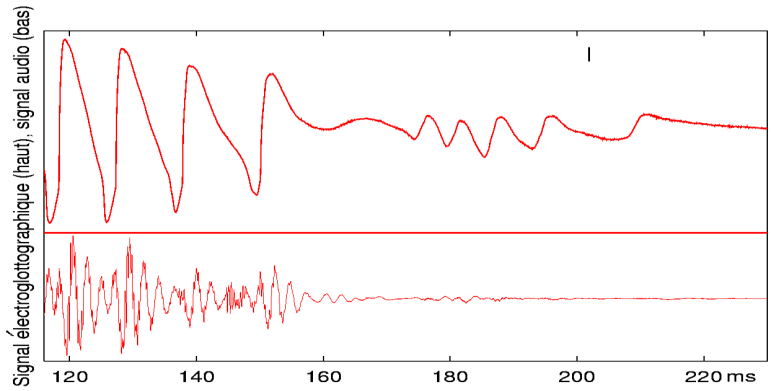


Fig. 3.65. Exemple d'un train de cycles glottiques de faible amplitude au ton B2. Syllabe /em/, locuteur M1.

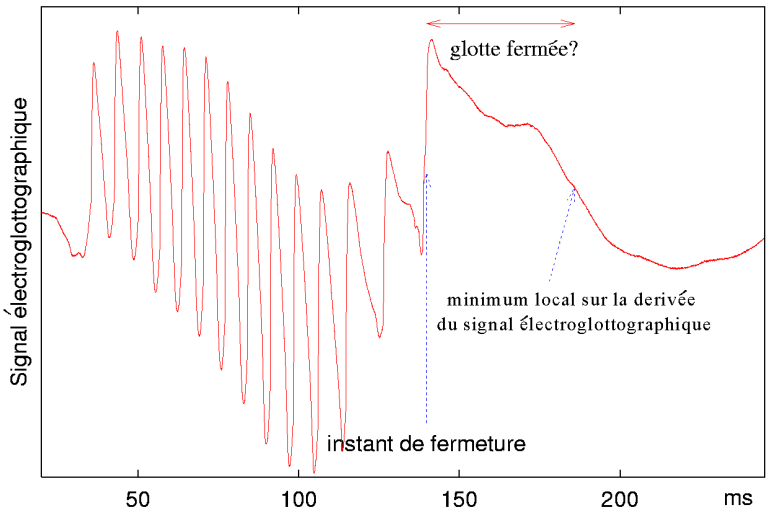


Fig. 3.66. Illustration de l'impossibilité d'estimer la durée de la phase fermée après la dernière fermeture en fin de voisement. Ton B2. Locuteur M1.

Fig. 3.67. Locuteur M1, ton B2 (=ton comportant une constriction glottale)

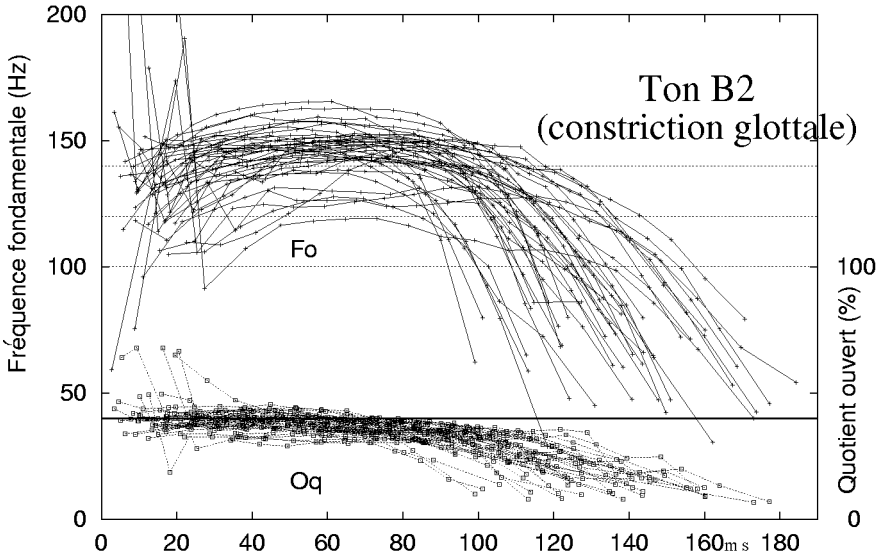


Fig. 3.67a. Courbes brutes de 42 items au ton B2, locuteur M1, lecture soignée.

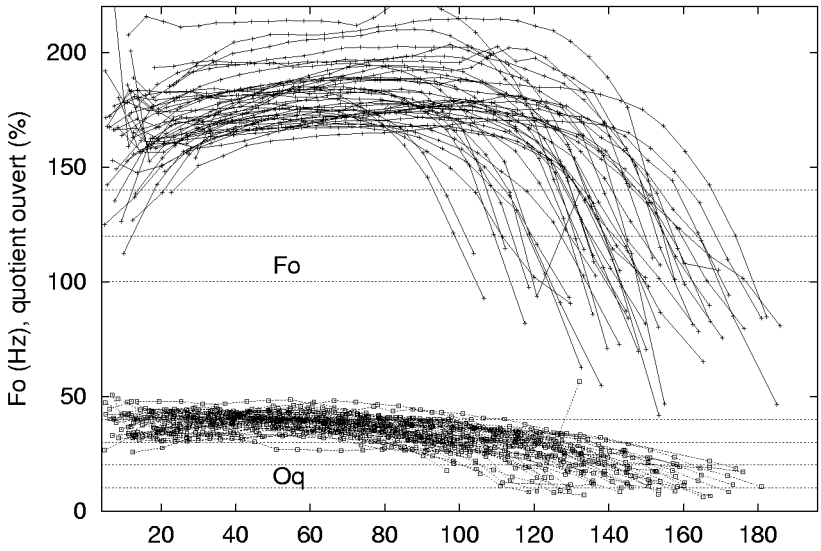


Fig. 3.67b. Courbes brutes de 42 items au ton B2, locuteur M1, lecture insistante.

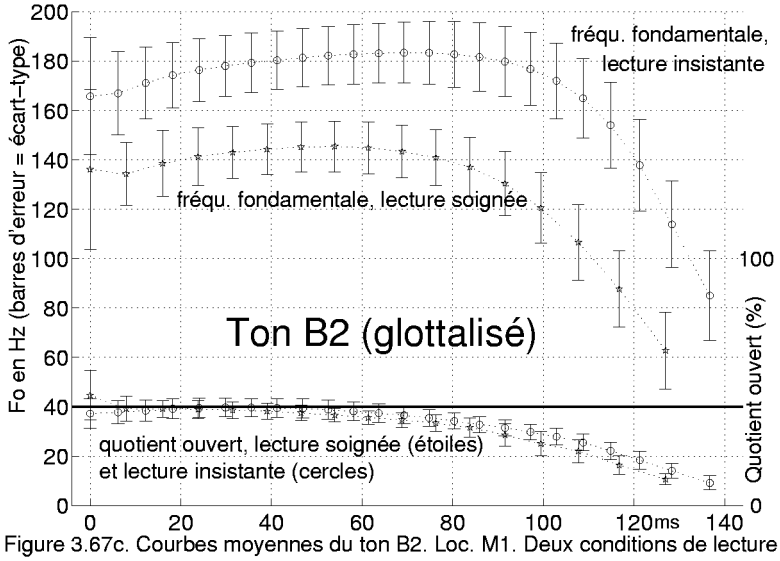


Figure 3.67c. Courbes moyennes du ton B2. Loc. M1. Deux conditions de lecture

Fig. 3.68. Locuteur vietnamien M1, ton D2 (=sans constriction glottale)

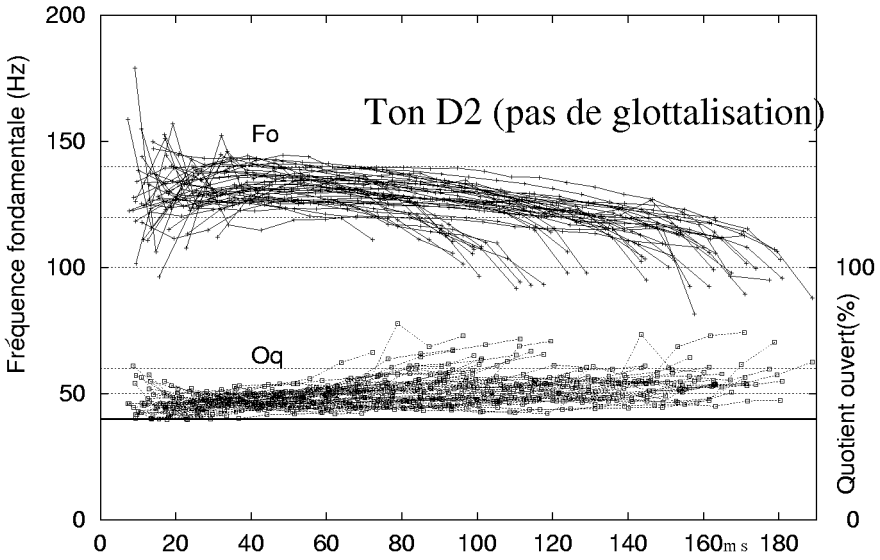


Fig. 3.68a. Courbes brutes de 42 items au ton D2, locuteur M1, lecture soignée.

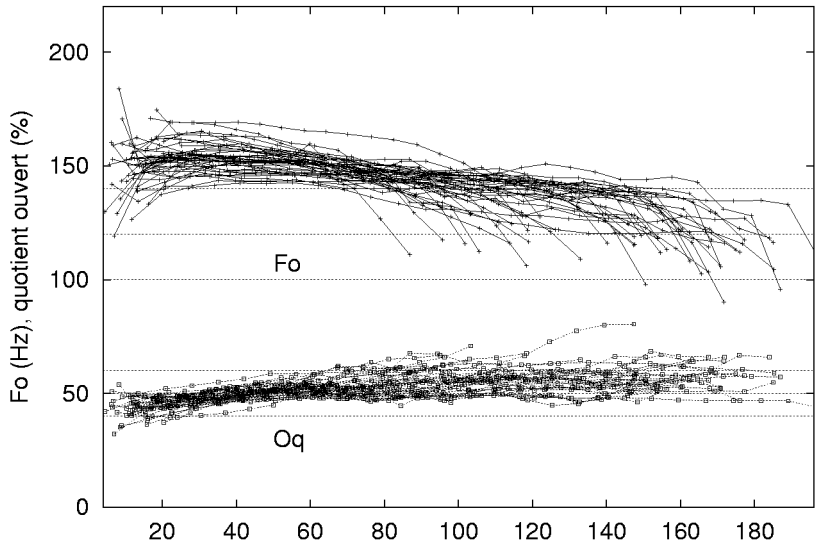


Figure 3.68b. Courbes brutes de 42 items au ton D2, locuteur M1, lecture insistante. Temps en ms.

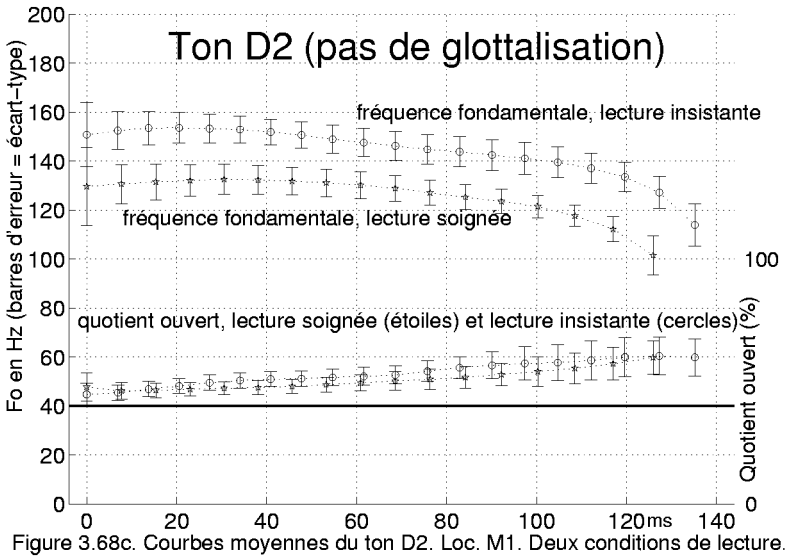


Figure 3.68c. Courbes moyennes du ton D2. Loc. M1. Deux conditions de lecture.

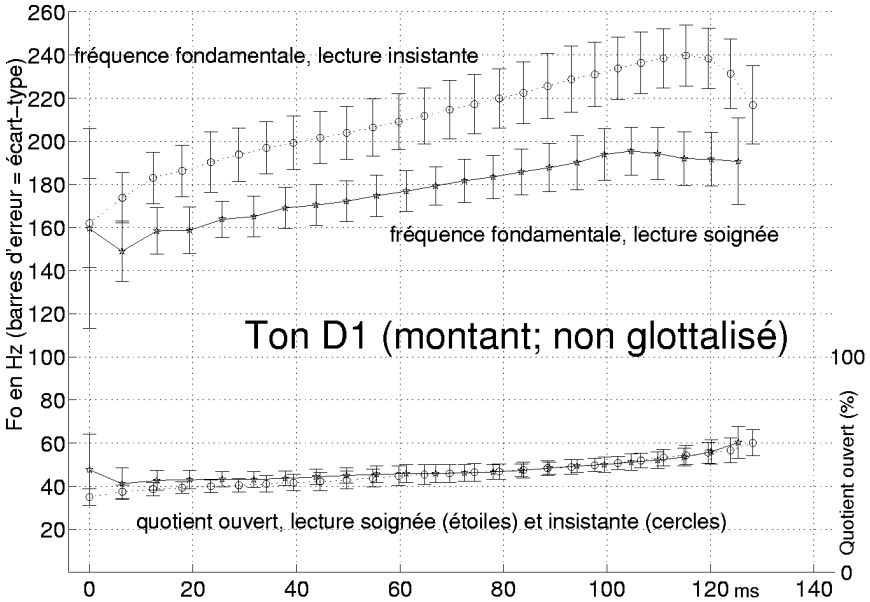


Fig. 3.69. Courbes moyennes du ton D1. Locuteur M1. Deux conditions de lecture.

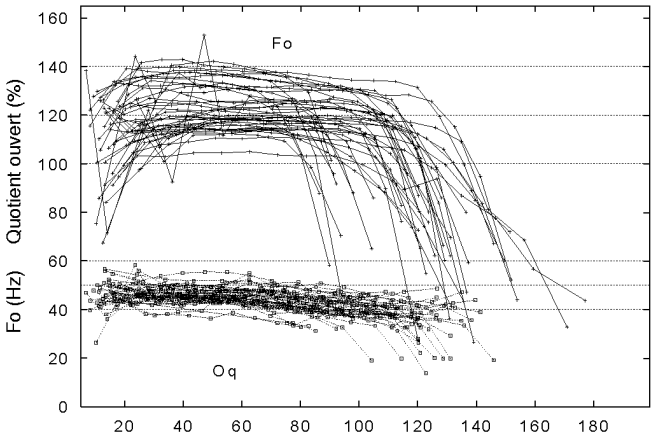


Figure 3.70a.
Courbes brutes de 42 items au ton B2, loc. M2, lecture soignée. Temps en ms.

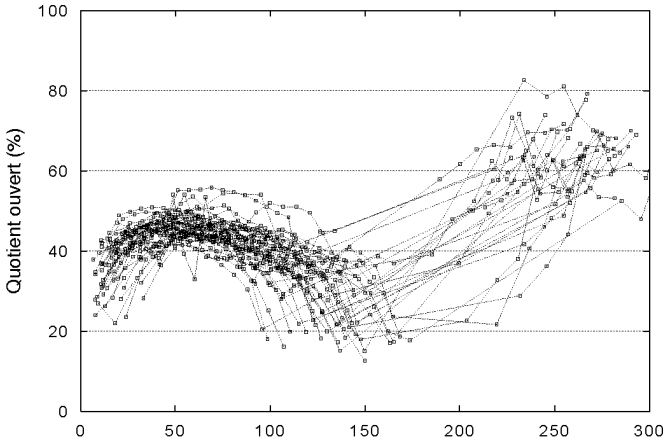


Figure 3.70c. Courbes brutes de 42 items au ton B2, loc. M2, lecture insistante. Temps en ms.

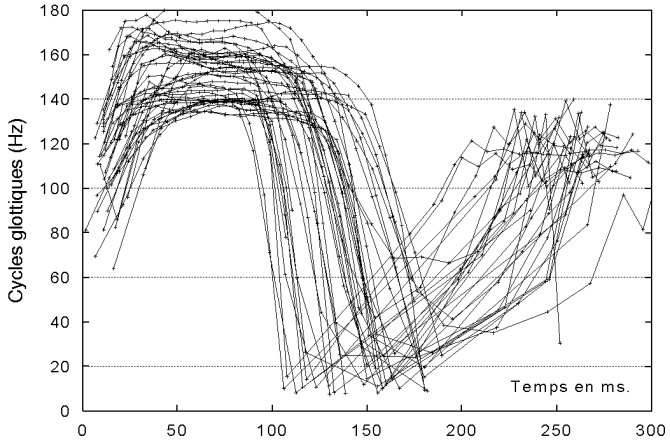


Fig. 3.70b. Courbes brutes de 42 items au ton B2, loc. M2, lecture insistante.
Des cycles glottiques dont le F0 est dans la plage moyenne succèdent à la constriction

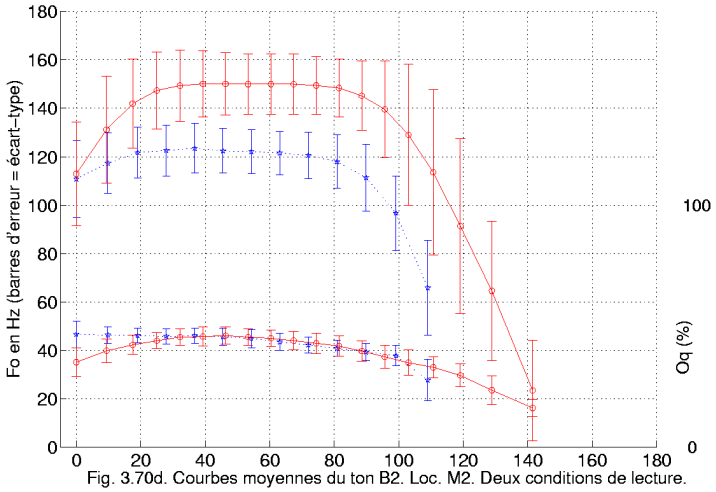


Fig. 3.70d. Courbes moyennes du ton B2. Loc. M2. Deux conditions de lecture.

Locuteur vietnamien M2

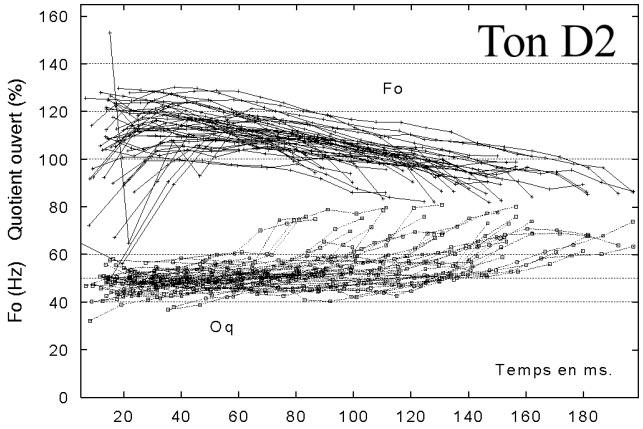


Fig. 3.71a. Courbes brutes de 42 items au ton D2, loc. M2, lecture soignée.

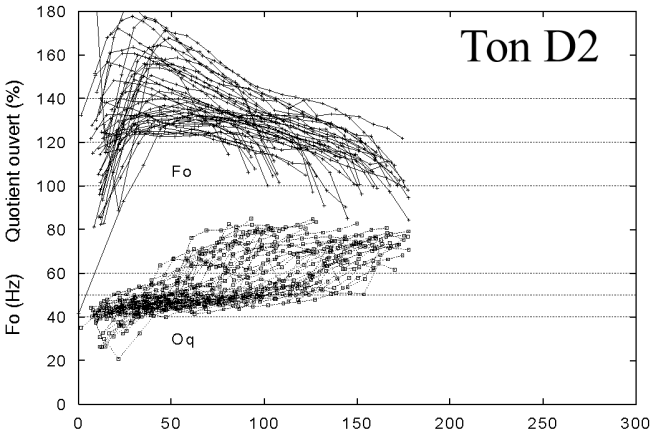


Fig. 3.71b. Courbes brutes de 42 items au ton D2, loc. M2, lecture insistante. Temps en ms.

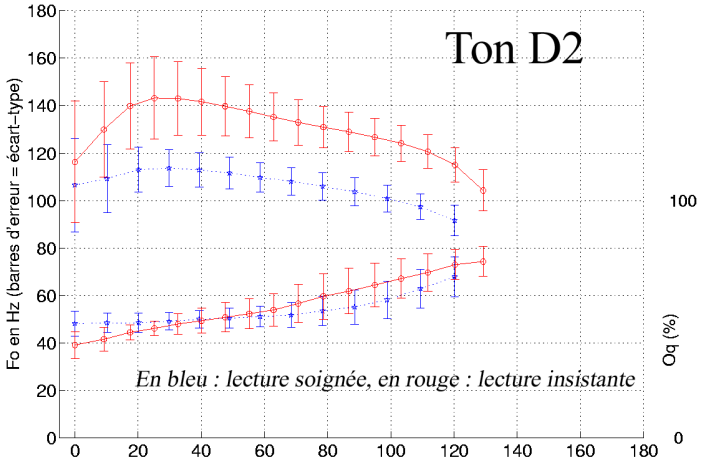


Fig. 3.71c. Courbes moyennes du ton D2. Loc. M2. Deux conditions. Temps en ms.

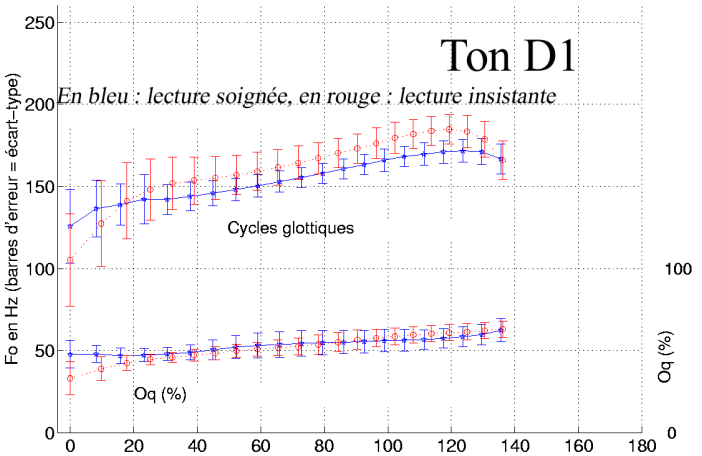


Fig. 3.71d. Courbes moyennes du ton D1. Locuteur M2. Conditions C et E. Temps en ms.

Locuteur vietnamien M3

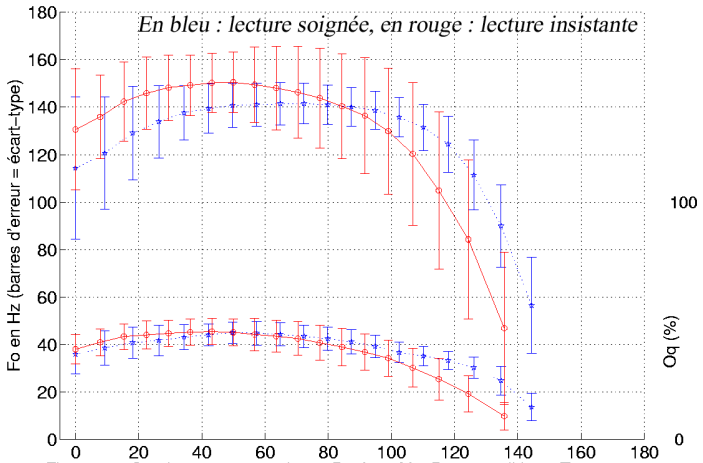


Fig. 3.72a. Courbes moyennes du ton B2. Loc. M3. Deux conditions. Temps en ms.

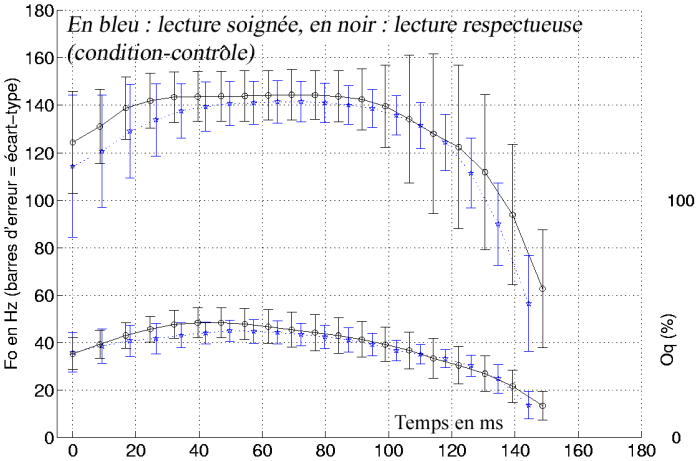


Fig. 3.72b. Courbes moyennes du ton B2. Locuteur M3. Condition soignée et respectueuse (condition-contrôle).

Locuteur vietnamien M3

Ton D2

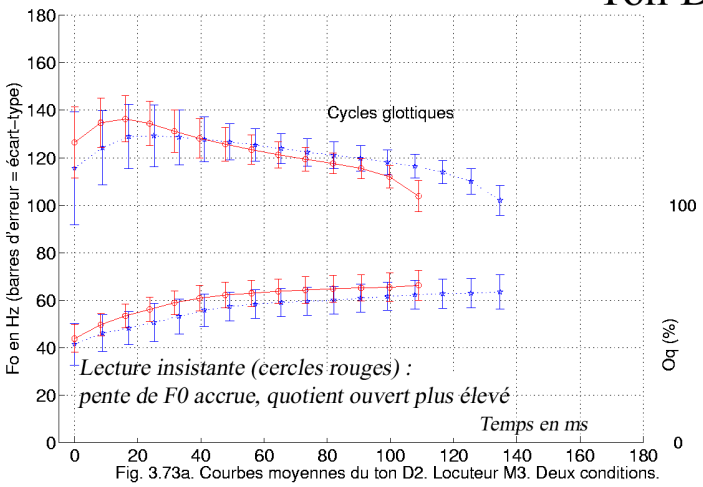
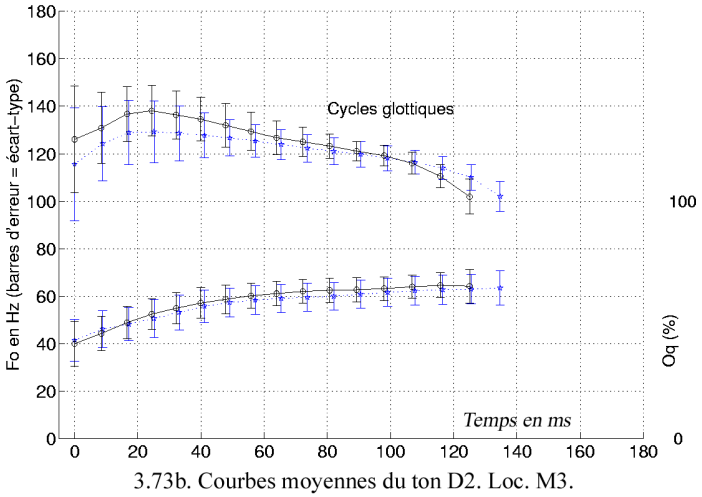
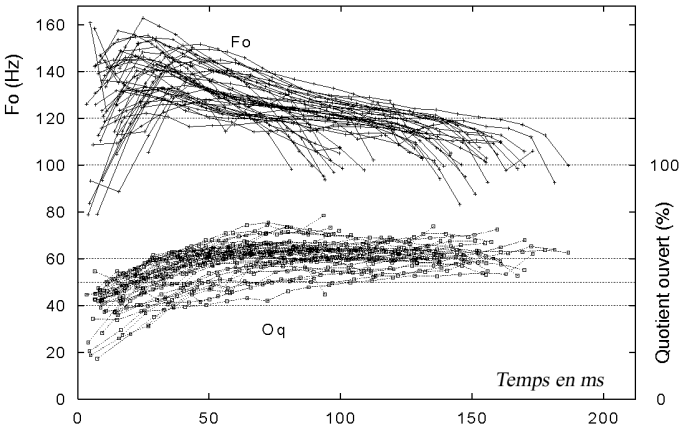


Fig. 3.73a. Courbes moyennes du ton D2. Locuteur M3. Deux conditions.



3.73b. Courbes moyennes du ton D2. Loc. M3.
Lecture soignée, lecture respectueuse (condition-contrôle).



3.73c. Courbes brutes de 42 items au ton D2, locuteur M3,
lecture respectueuse (condition-contrôle).

Ton D1

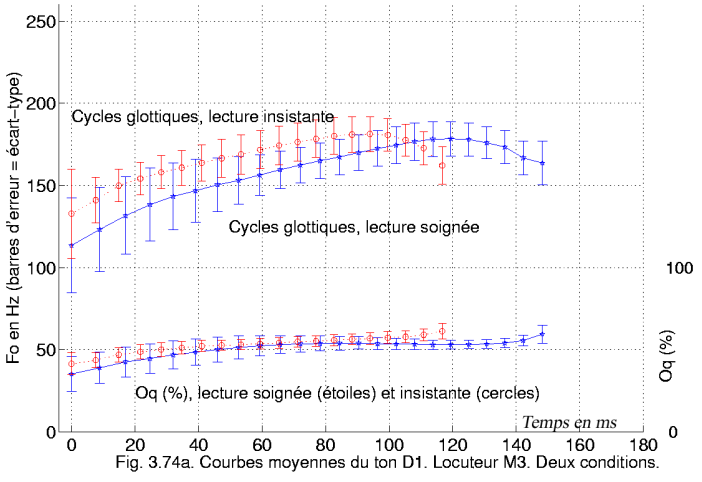
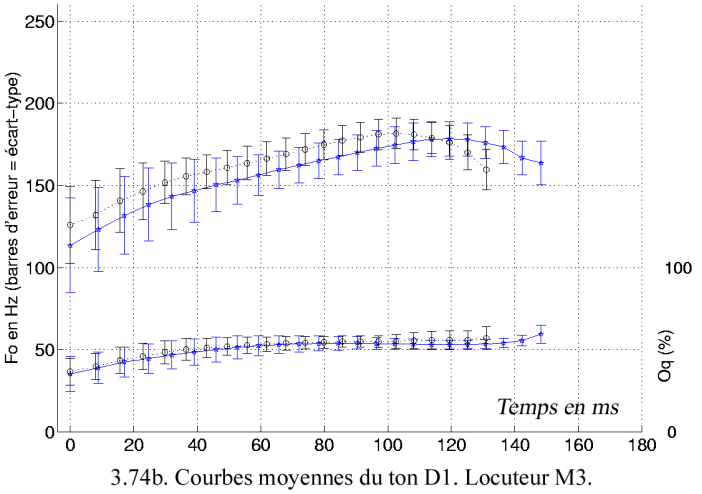


Fig. 3.74a. Courbes moyennes du ton D1. Locuteur M3. Deux conditions.



3.74b. Courbes moyennes du ton D1. Locuteur M3.
Lecture soignée, lecture respectueuse (condition-contrôle).

Locutrice vietnamienne F1

Rappel : lecture soignée en bleu, lecture insistante en rouge.

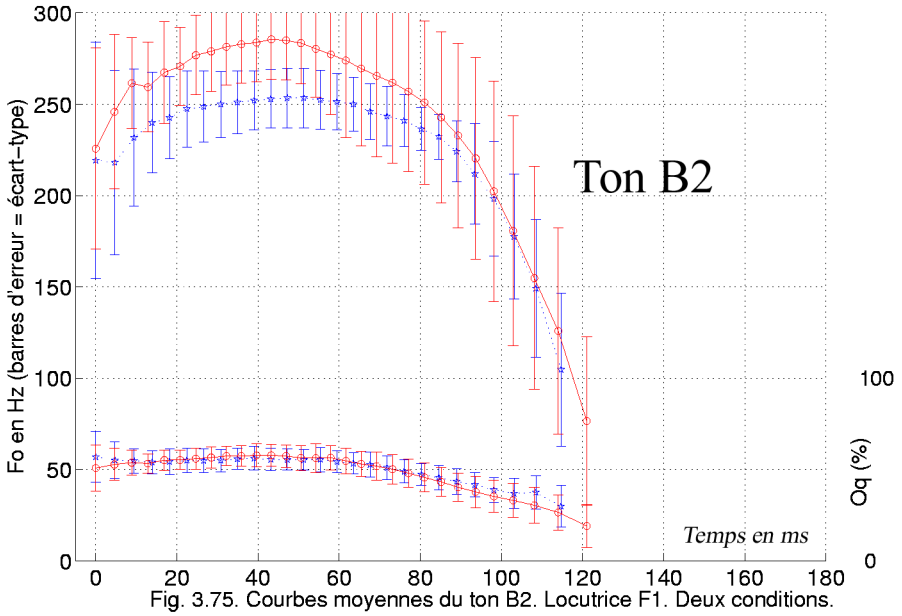


Fig. 3.75. Courbes moyennes du ton B2. Locutrice F1. Deux conditions.

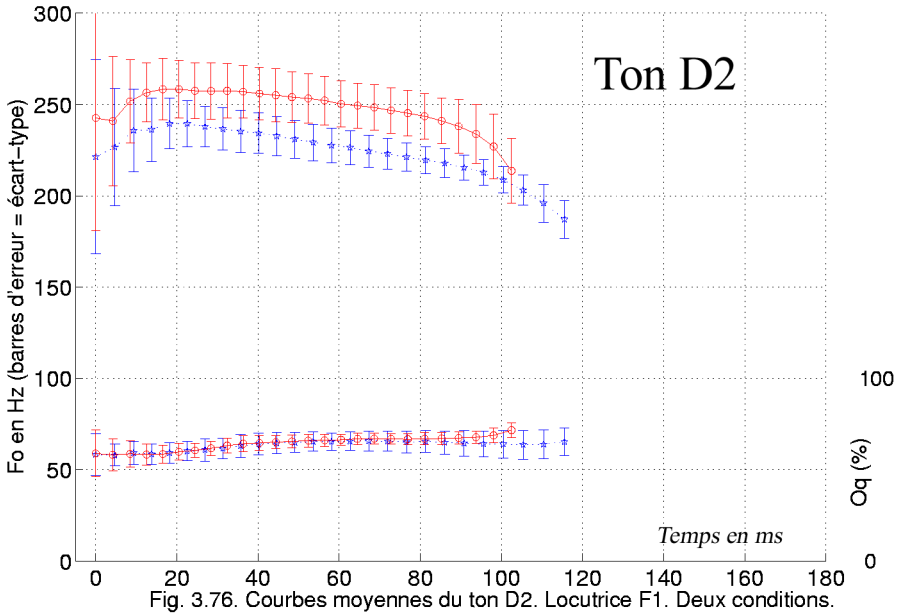


Fig. 3.76. Courbes moyennes du ton D2. Locutrice F1. Deux conditions.

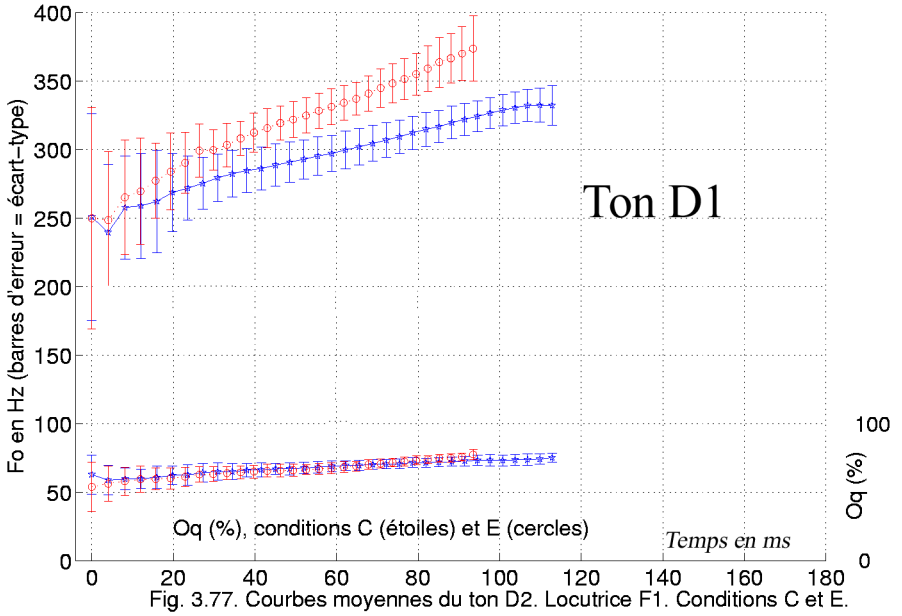


Fig. 3.77. Courbes moyennes du ton D2. Locutrice F1. Conditions C et E.

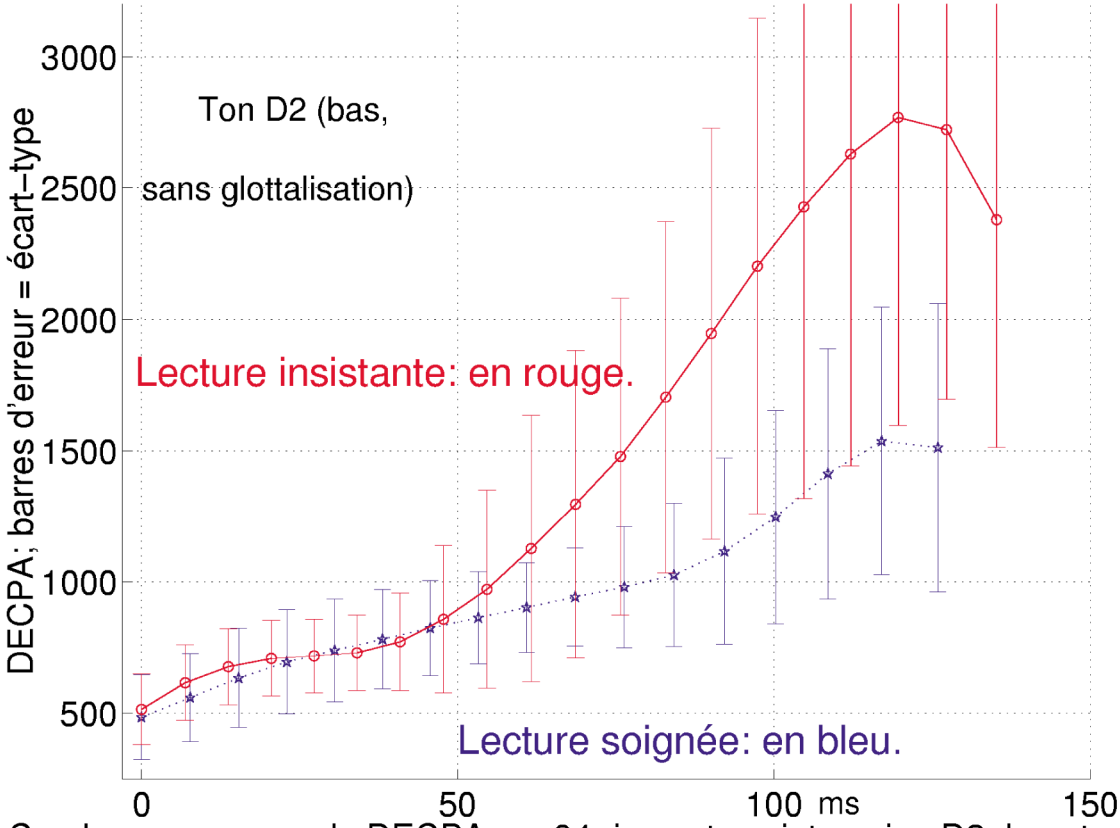


Fig. 3.78. Courbes moyennes de DECPA sur 84 rimes, ton vietnamien D2. Locuteur M1.

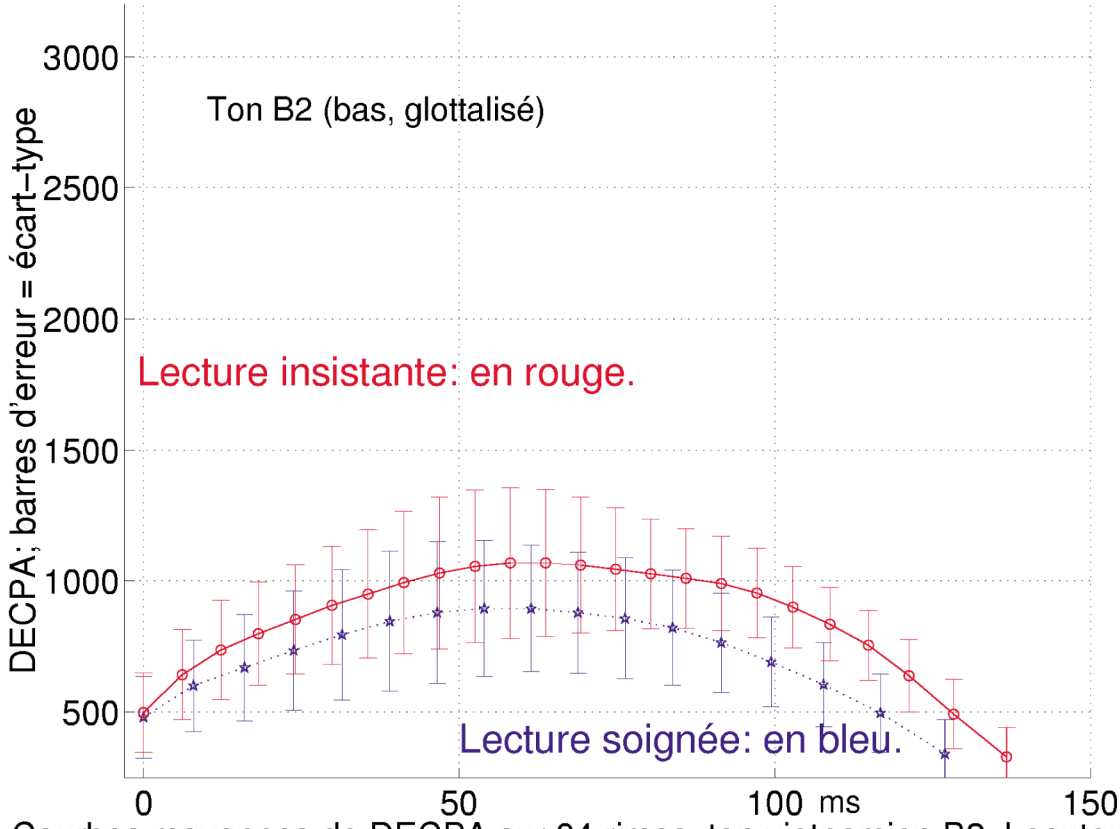


Fig. 3.79. Courbes moyennes de DECPA sur 84 rimes, ton vietnamien B2. Locuteur M1.

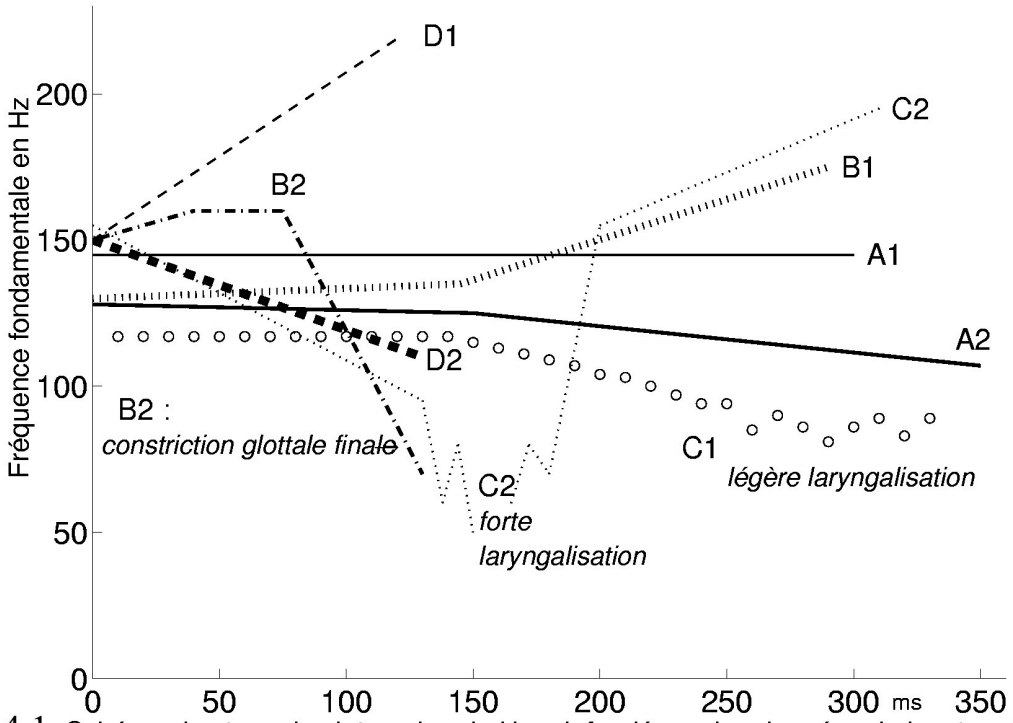
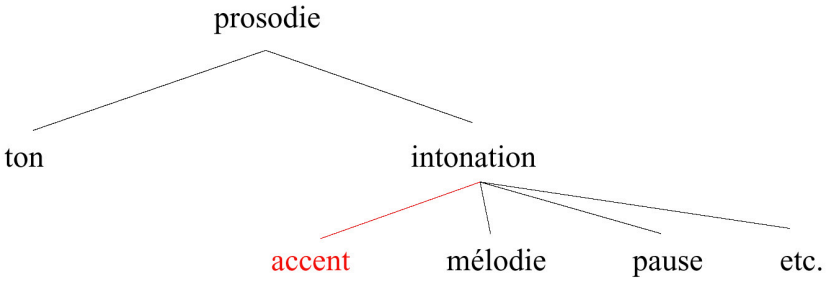
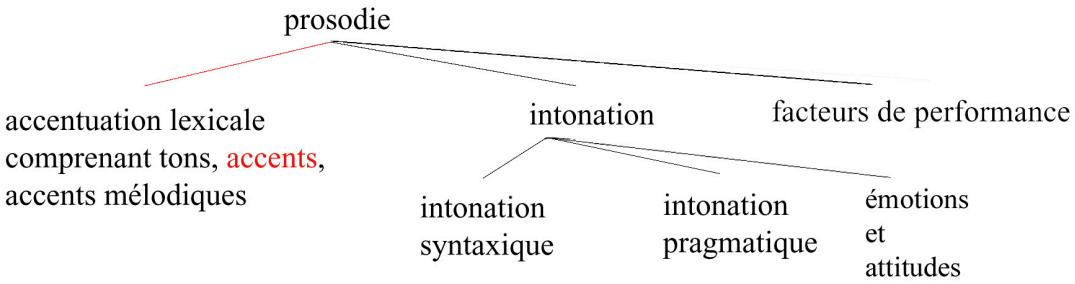


Fig. 4.1. Schéma des tons du vietnamien de Hanoi, fondé sur les données du locuteur M1.



a. Représentation des composantes de la prosodie par Hoa 1980 : *L'Accentuation en pékinois.* L'accent figure parmi les composantes de l'intonation.



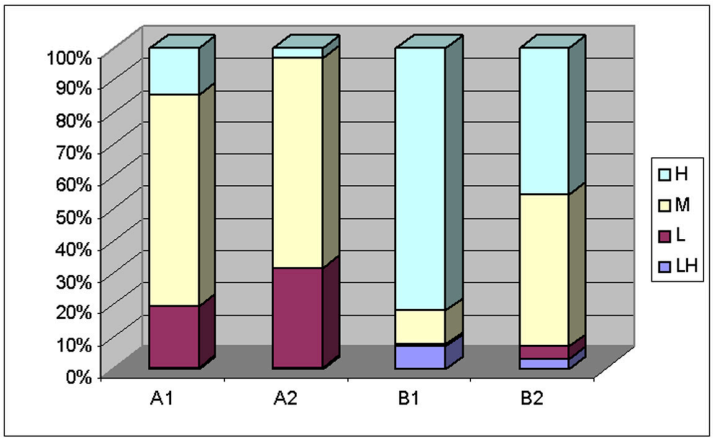
b. Représentation des composantes de la prosodie par Vaissière 2004. La notion d'accent est limitée au domaine de l'accentuation lexicale (ensemble des phénomènes "suprasegmentaux" qui contribuent à distinguer les mots dans le lexique).

Fig. 4.2. Place de l'accentuation dans la prosodie chez M. Hoa (haut) et J. Vaissière (bas).

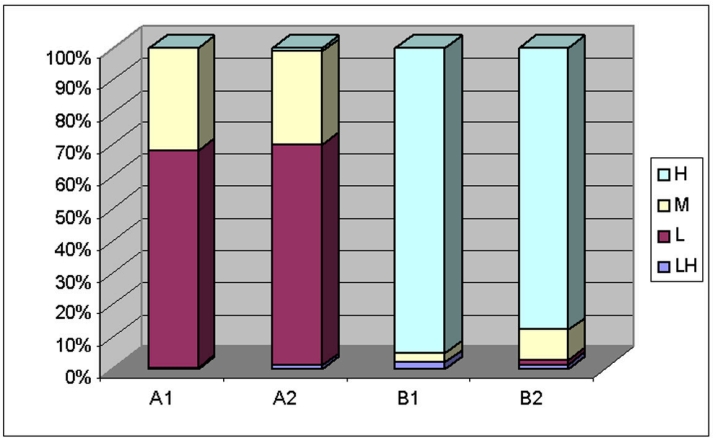
Graphiques

Une table des graphiques est proposée en page 520.

Graphiques du Chapitre II

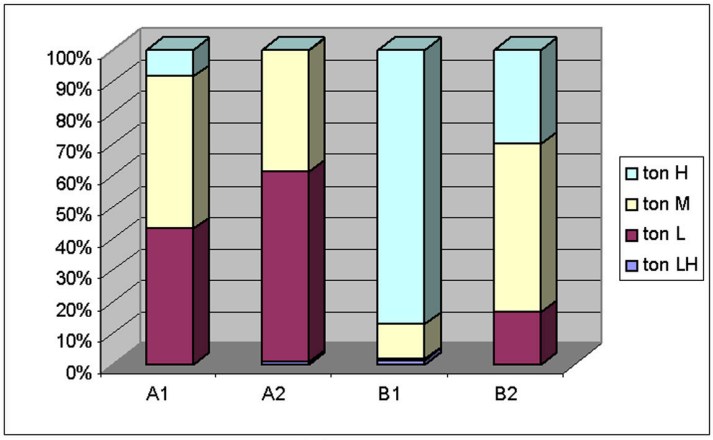


Graphique 2.1.

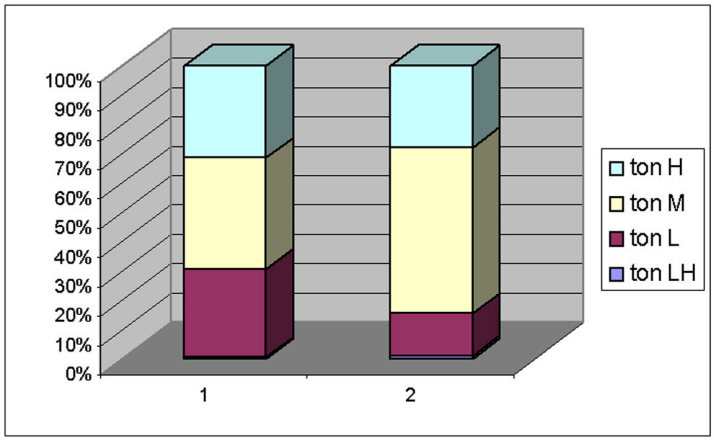


Graphique 2.2.

Graphiques 2.1 et 2.2. Résultats du test d'identification du ton des syllabes isolées extraites des séquences réduplicquées A1 B1 A2 B2. 40 stimuli (signal électroglottographique), 15 auditeurs. A gauche (2.1): données produites par le locuteur M1; à droite (2.2): données produites par le locuteur M4.



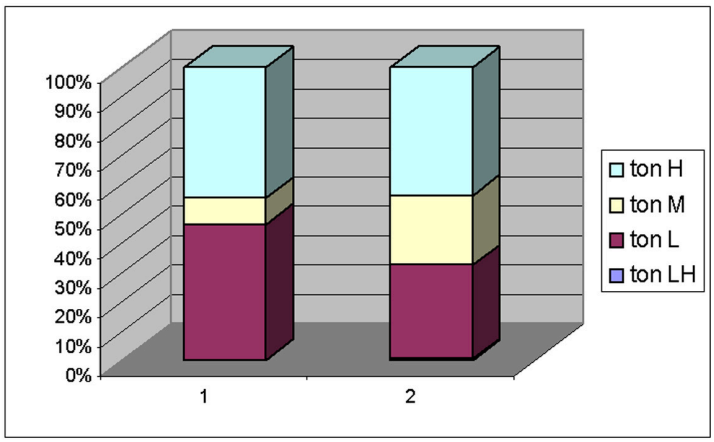
Graphique 2.3.



Graphique 2.4.

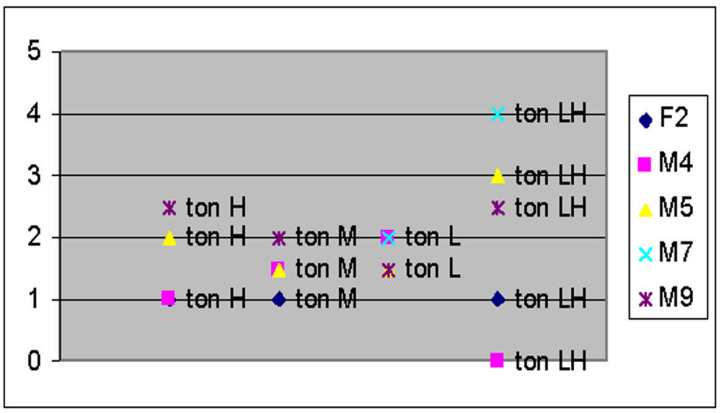
Graphique 2.3. Identification du ton des syllabes extraites des séquences réduplicquées A1 B1 A2 B2 lorsqu'elles figurent en première position d'un stimulus disyllabique. 140 stimuli, 14 auditeurs. Données produites par le locuteur M1.

(légende commune avec 2.5 ; voir ci-dessous)



Graphique 2.5.

Graphiques 2.4 et 2.5. 2.4 Identifications du ton des syllabes extraites des séquences réduplicquées A1 B1 A2 B2 en fonction de la position dans le composé disyllabique utilisé comme stimulus. 140 stimuli, 14 auditeurs ; données produites par le locuteur M1 (graphique 2.1) et par M2 (graphique 2.2).



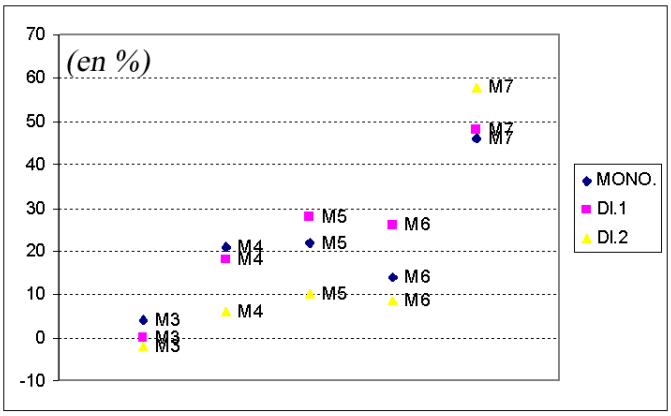
Graphique 2.6. Approximation en tons musicaux de la déclinaison à l'échelle de l'énoncé dans les énoncés naxi dont toutes les syllabes portent un même ton lexical. Chaque point correspond à la moyenne de plusieurs énoncés (voir détails sur les figures 2.13 à 2.27). Chaque couleur correspond à un locuteur.

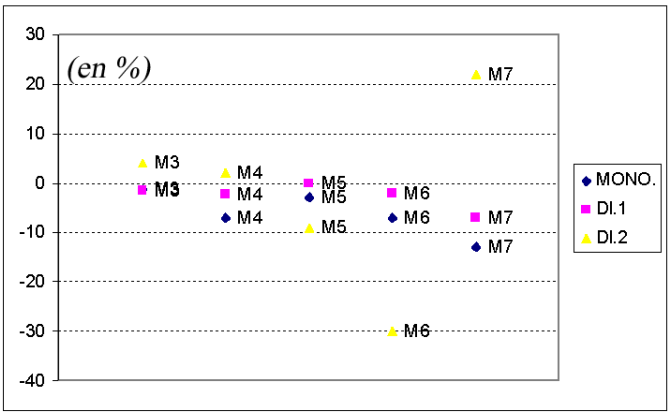
Graphiques 3.1 à 3.9 : différences observées entre conditions de lecture dans les trois langues

Fréquence fondamentale

Quotient ouvert

anglais

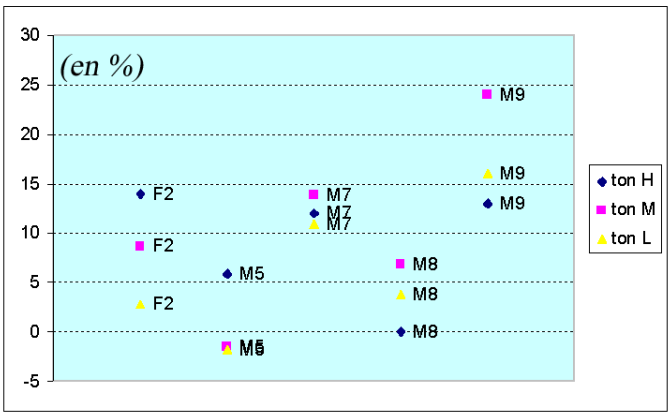


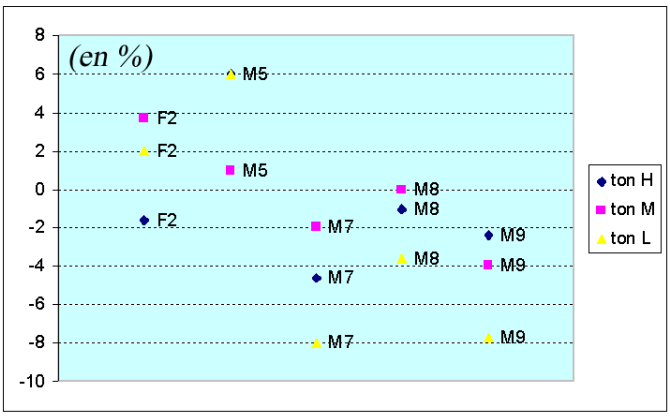


Graphique 3.1. Différence moyenne de F0 entre conditions de lecture, en fonction du locuteur et du type de syllabe (MONO : monosyllabes, DI. 1 et 2 : 1e et 2e syllabe de disyllabe, respectivement).

Graphique 3.2. Différence moyenne de quotient ouvert entre conditions de lecture, en fonction du locuteur et du type de syllabe (MONO : monosyllabes, DI. 1 et 2 : 1e et 2e syllabe de disyllabe, respectivement).

naxi

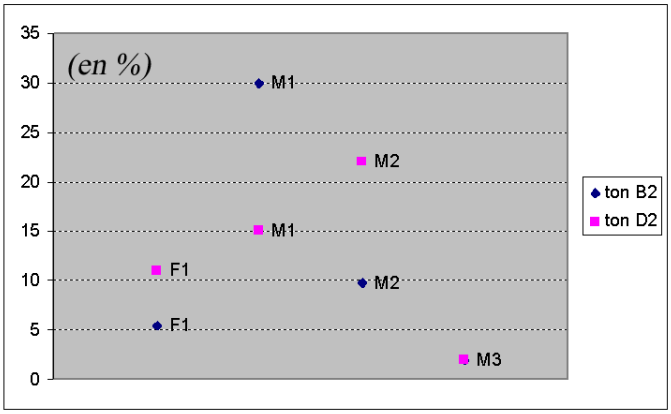


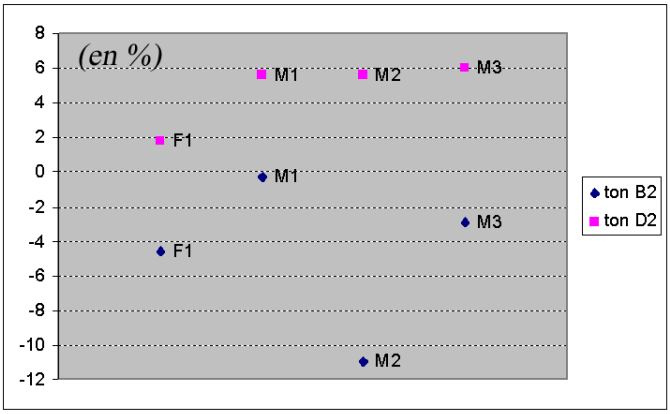


Graphique 3.4. Différence moyenne de F0 entre conditions de lecture, en fonction du locuteur et du ton de la syllabe (H : ton haut, M : ton moyen, L : ton bas).

Graphique 3.5. Différence moyenne de quotient ouvert entre conditions de lecture, en fonction du locuteur et du ton de la syllabe (H : ton haut, M : ton moyen, L : ton bas).

vietnamien

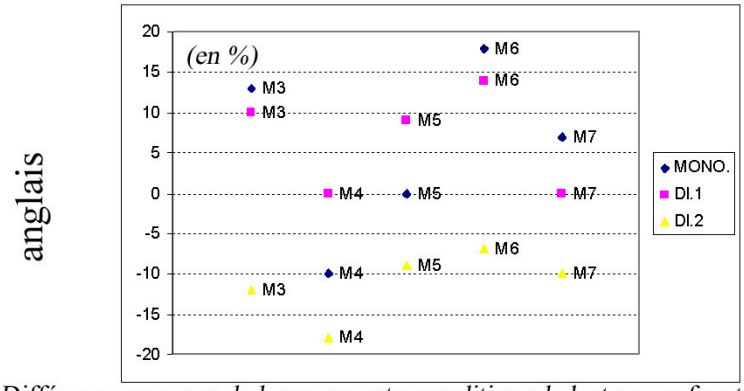




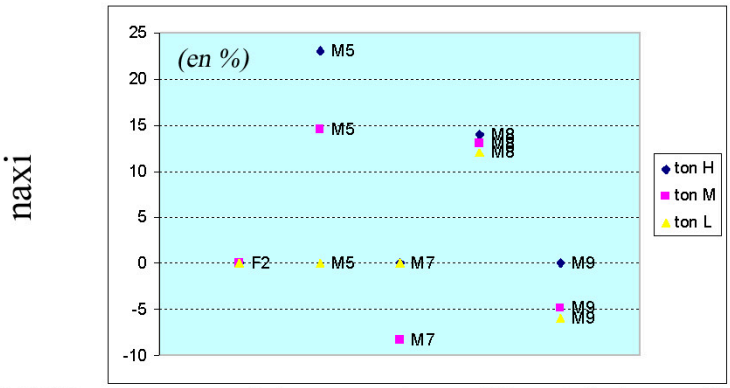
Graphique 3.7. Différence moyenne de F0 entre conditions de lecture, en fonction du locuteur et du ton de la syllabe (B2 : ton glottalisé, D2 : ton non glottalisé).

Graphique 3.8. Différence moyenne de quotient ouvert entre conditions de lecture, en fonction du locuteur et du ton de la syllabe (B2 : ton glottalisé, D2 : ton non glottalisé)

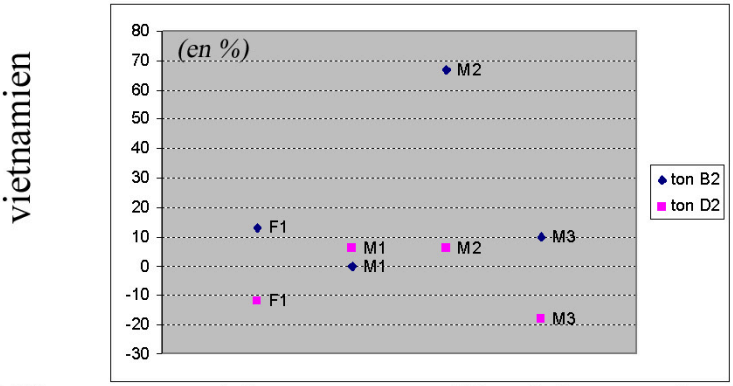
Graphiques 3.1 à 3.9 : différences observées entre conditions de lecture dans les trois langues



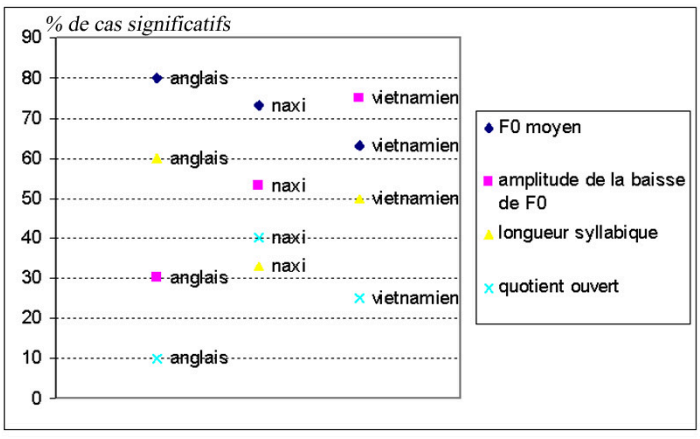
Graphique 3.3. Différence moyenne de longueur entre conditions de lecture, en fonction du locuteur et du type de syllabe (MON DI. 1 et 2 : 1e et 2e syllabe de disyllabe, respectivement).



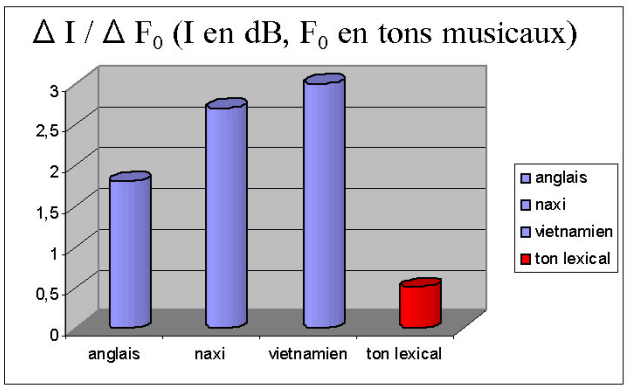
Graphique 3.6. Différence moyenne de longueur entre conditions de lecture, en fonction du locuteur et du ton de la syllabe



Graphique 3.9. Différence moyenne de longueur entre conditions de lecture, en fonction du locuteur et du ton de la syllabe (B2 : ton glottalisé, D2 : ton non glottalisé).



Graphique 3.10. Bilan simplifié des différences entre conditions de lecture dans les trois langues.



Graphique 3.11. Rapport entre variation de F0 et variation d'intensité acoustique globale : d'une condition de lecture à l'autre, dans les trois langues (en bleu), et entre tons lexicaux M et H en naxi (en rouge).

Figures des Annexes

Une table des figures des Annexes est proposée en page 521.

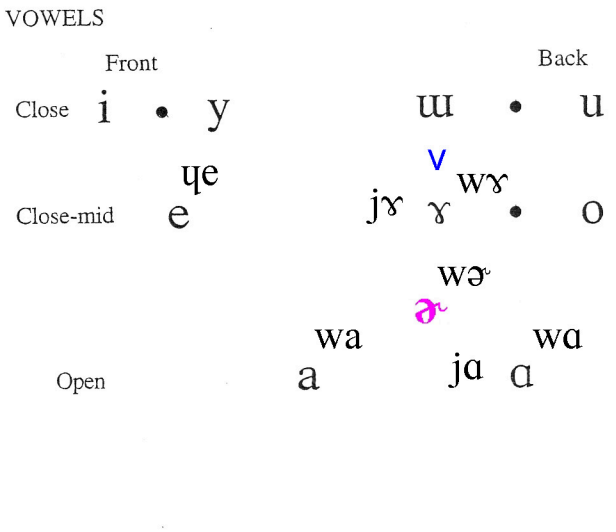


Figure 1.1. Les phonèmes vocaliques simples du naxi de AS. Trois degrés d'aperture, deux degrés d'antériorité, une opposition d'arrondissement ; un /v/ syllabique et une voyelle rhotique.

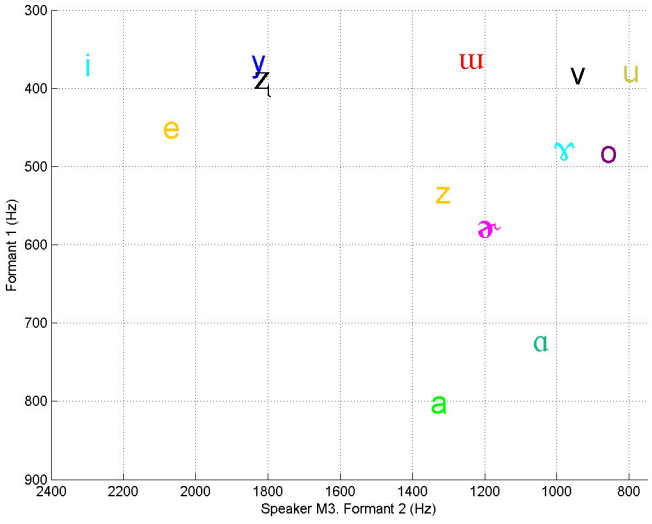


Figure 1.2. Estimation des fréquences des deux premiers formants des principales voyelles naxi. Locuteur M3.

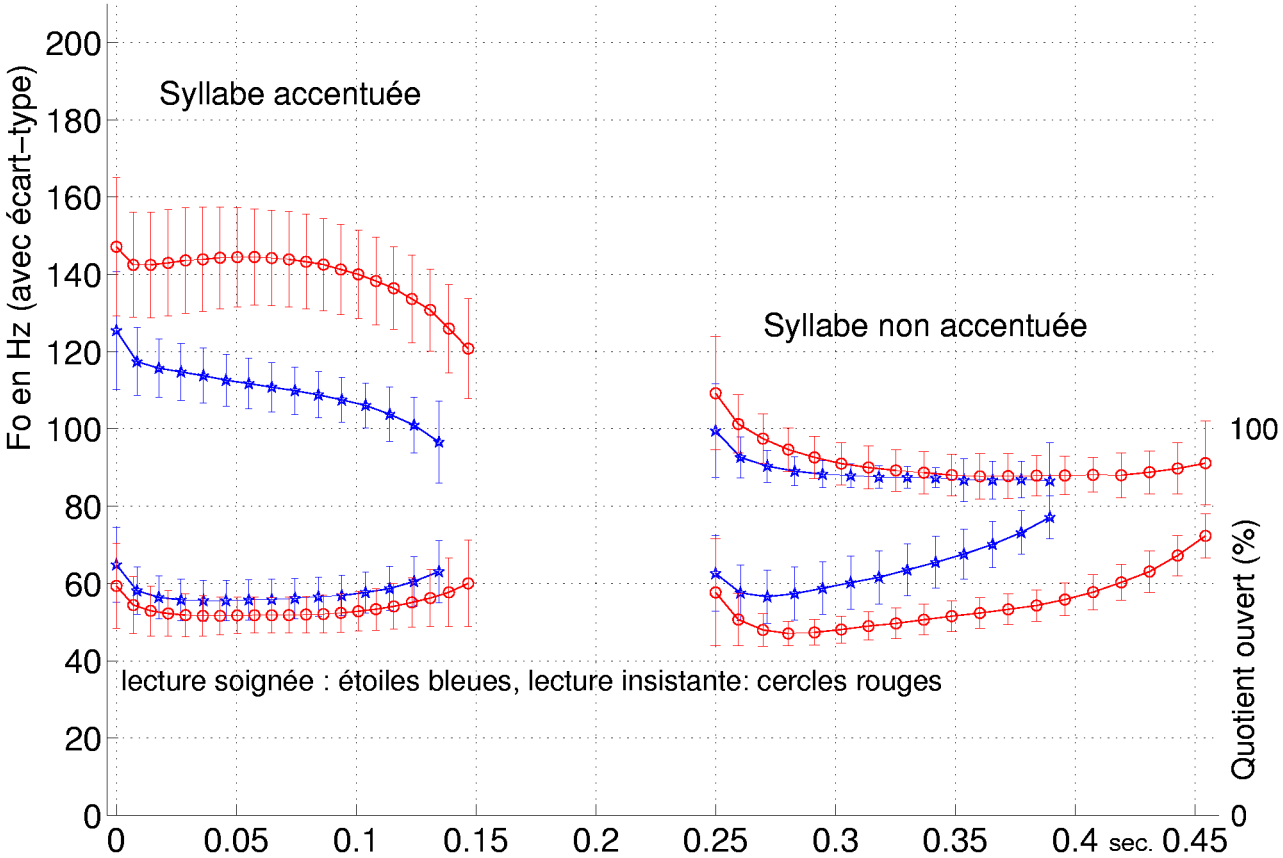


Fig. 2.1. Courbes moyennes de 317 disyllabes, locuteur anglais M1, lecture soignée et insistante.

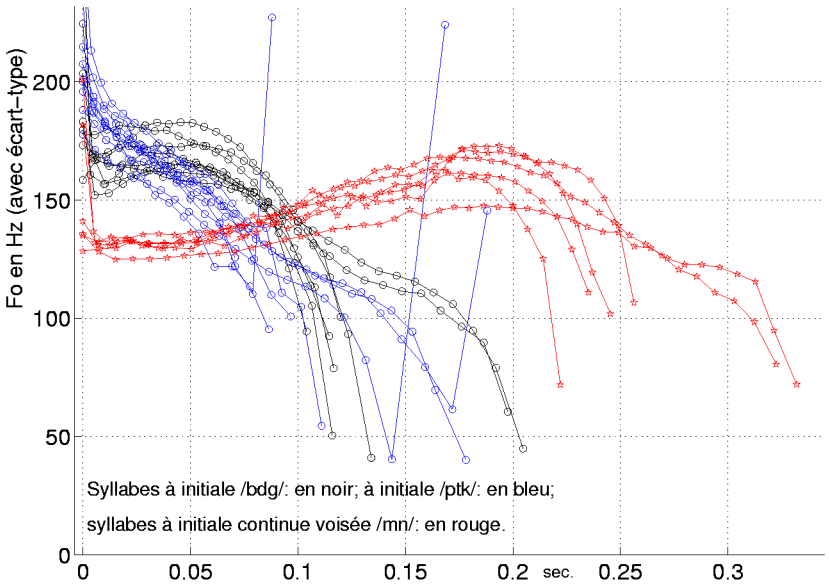


Fig. 2.2. Courbes brutes de 24 monosyllabes; langue: anglais; locuteur M4, lecture soignée.

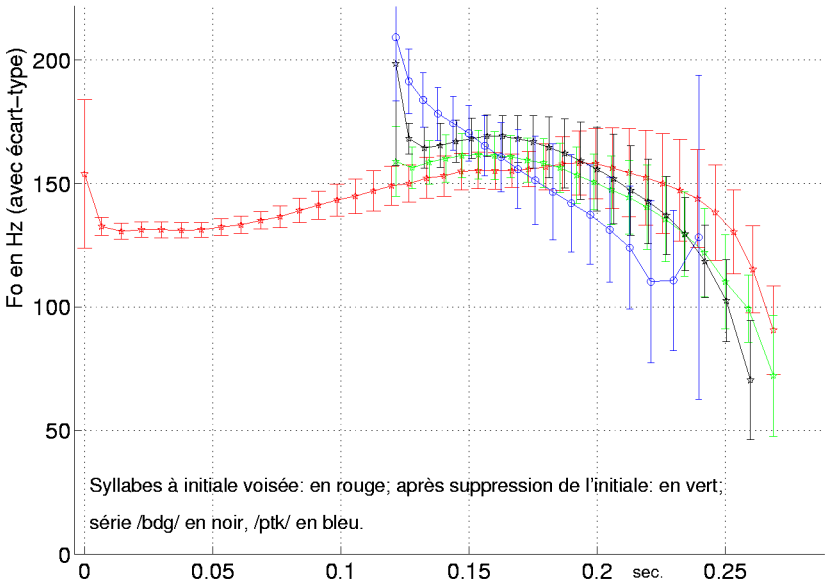


Fig. 2.3. Courbes moyennes de 24 monosyllabes; langue: anglais; locuteur M4, lecture soignée.

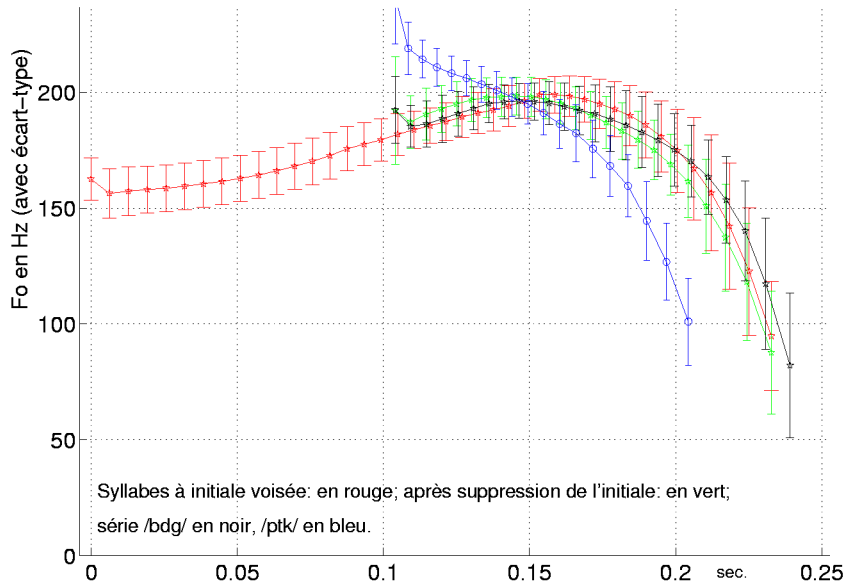


Fig. 2.4. Courbes moyennes de 24 monosyllabes; langue: anglais; locuteur M4, lecture insistante.

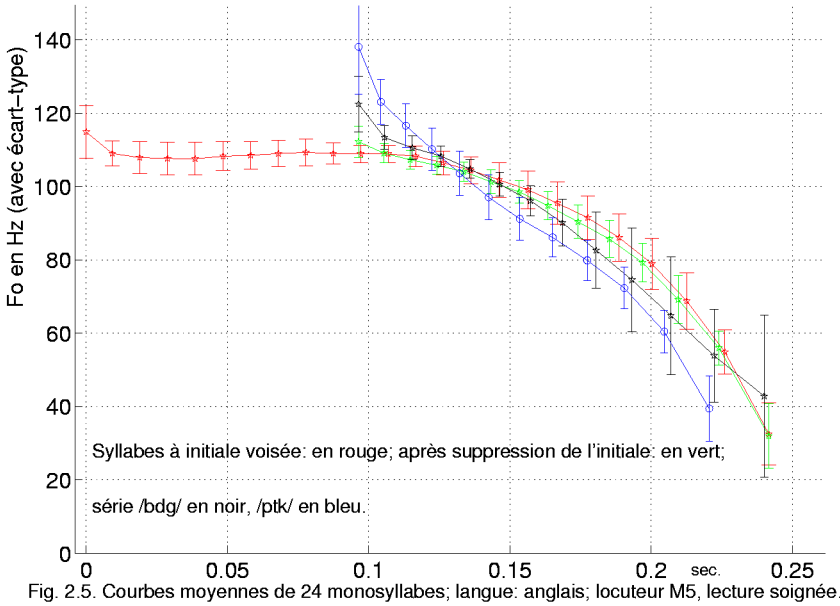


Fig. 2.5. Courbes moyennes de 24 monosyllabes; langue: anglais; locuteur M5, lecture soignée.

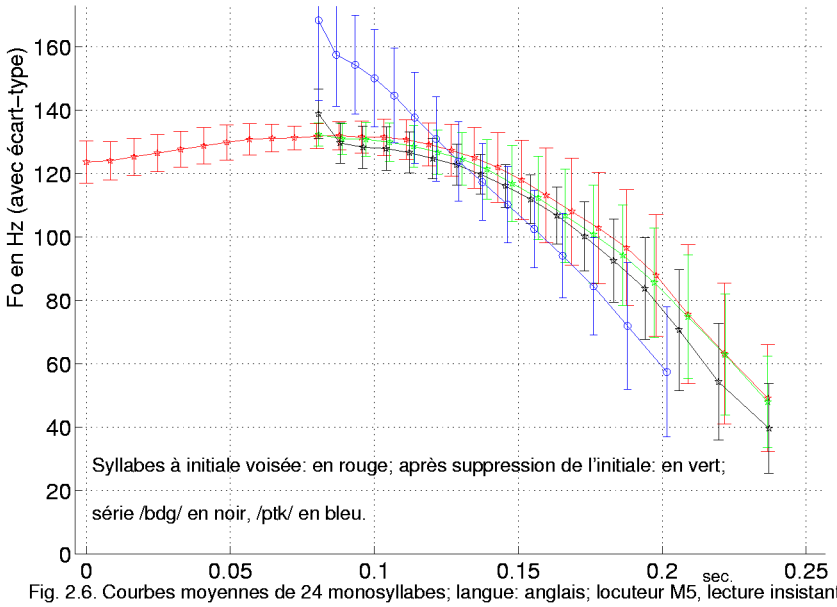


Fig. 2.6. Courbes moyennes de 24 monosyllabes; langue: anglais; locuteur M5, lecture insistante.

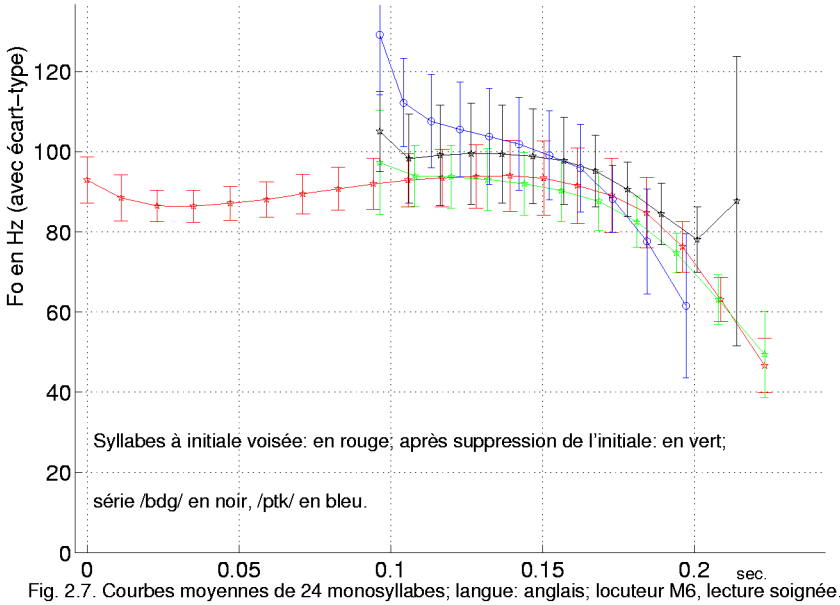


Fig. 2.7. Courbes moyennes de 24 monosyllabes; langue: anglais; locuteur M6, lecture soignée.

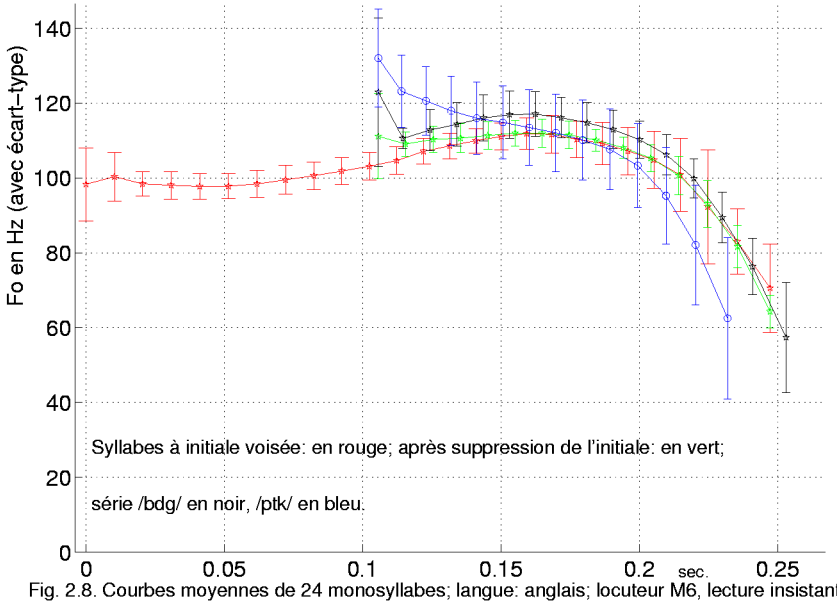


Fig. 2.8. Courbes moyennes de 24 monosyllabes; langue: anglais; locuteur M6, lecture insistante.

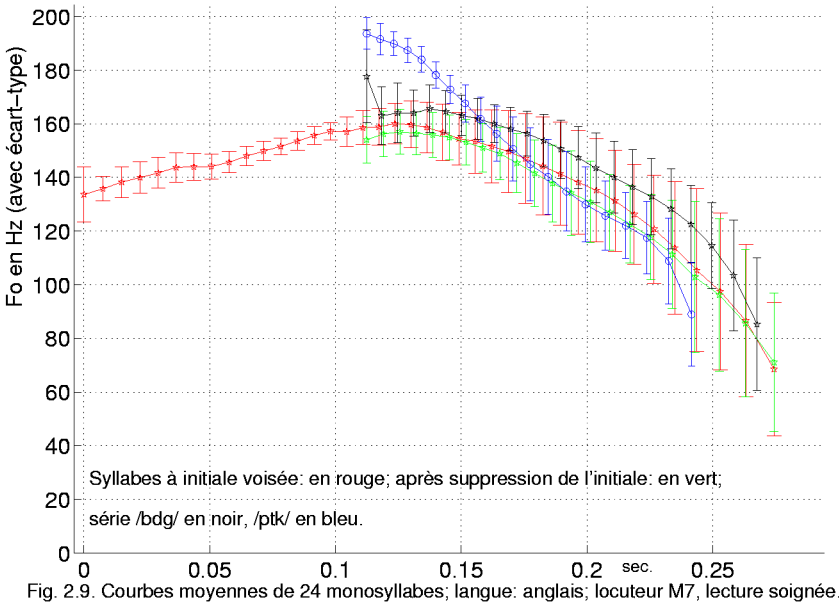


Fig. 2.9. Courbes moyennes de 24 monosyllabes; langue: anglais; locuteur M7, lecture soignée.

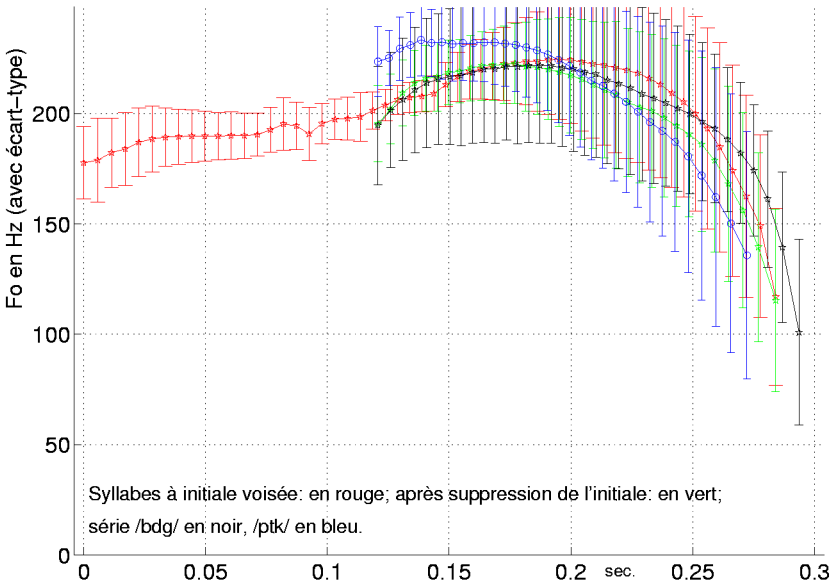


Fig. 2.10. Courbes moyennes de 24 monosyllabes; langue: anglais; locuteur M7, lecture soignée.

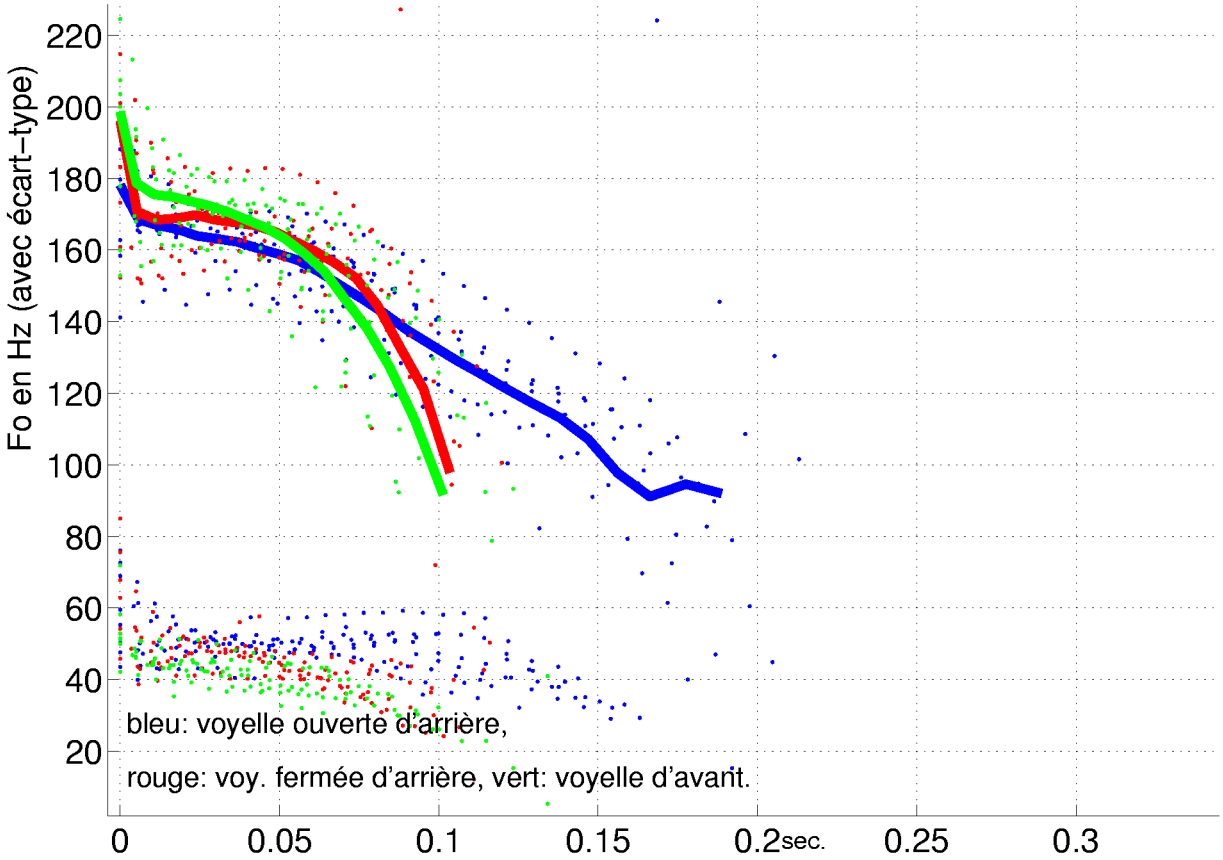


Fig. 2.11. Courbes moyennes (lignes) et brutes (points) de 24 monosyllabes; loc. M4, lecture soignée.

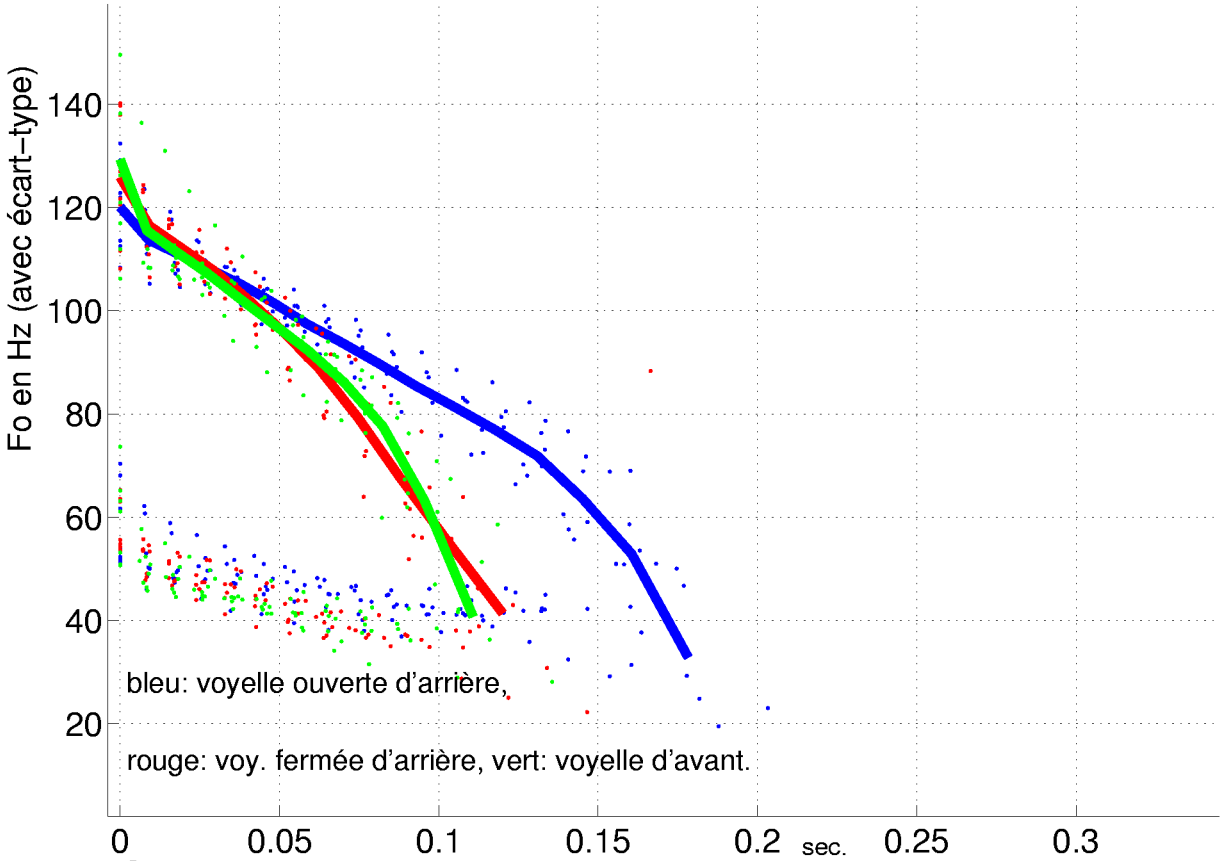


Fig. 2.12. Courbes moyennes (lignes) et brutes (points) de 24 monosyllabes; loc. M5, lecture soignée.

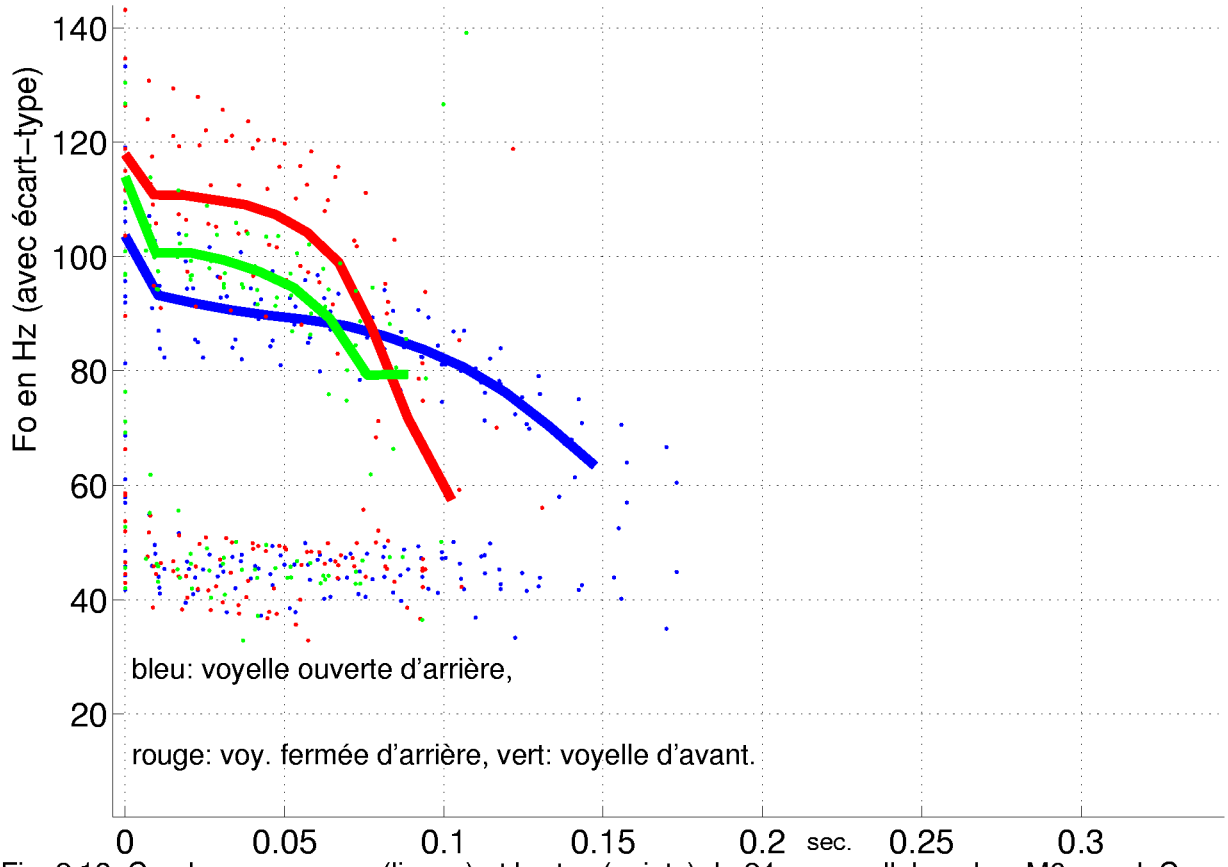


Fig. 2.13. Courbes moyennes (lignes) et brutes (points) de 24 monosyllabes; loc. M6, cond. C.

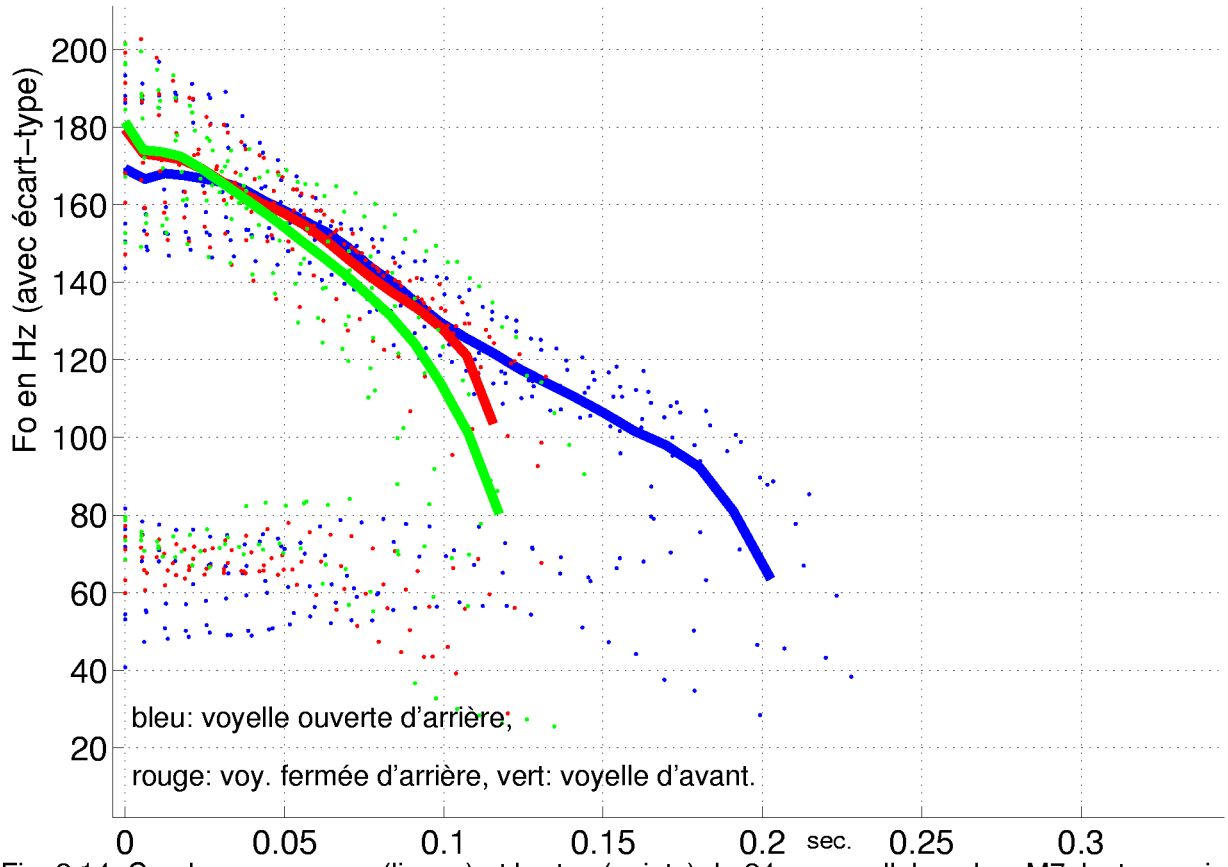


Fig. 2.14. Courbes moyennes (lignes) et brutes (points) de 24 monosyllabes; loc. M7, lecture soignée.

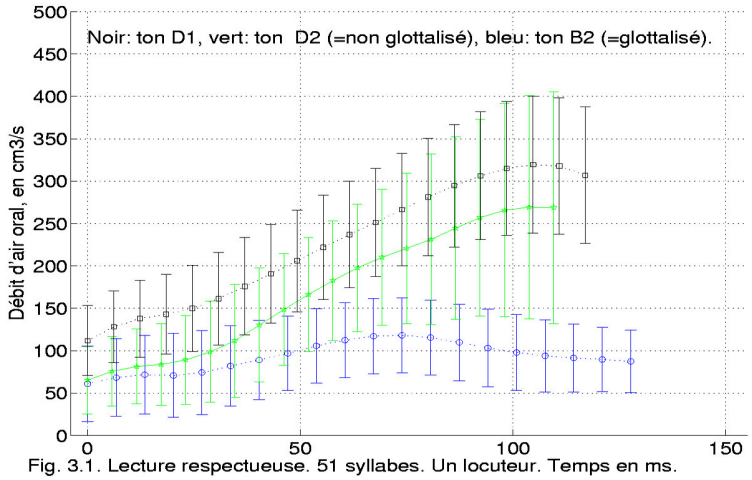


Fig. 3.1. Lecture respectueuse. 51 syllabes. Un locuteur. Temps en ms.

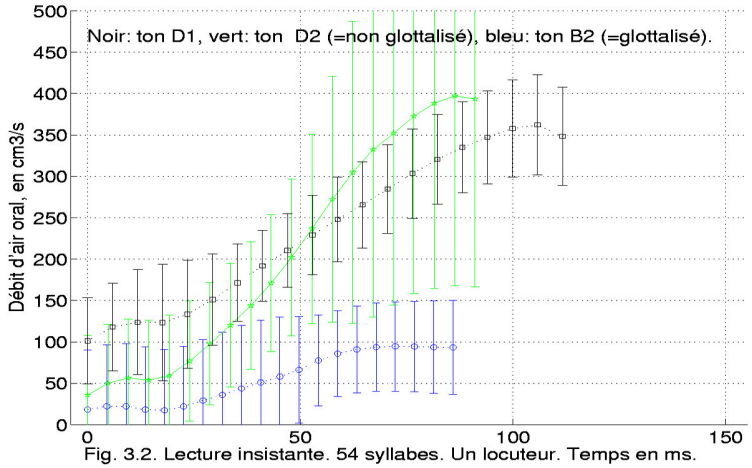


Fig. 3.2. Lecture insistante. 54 syllabes. Un locuteur. Temps en ms.

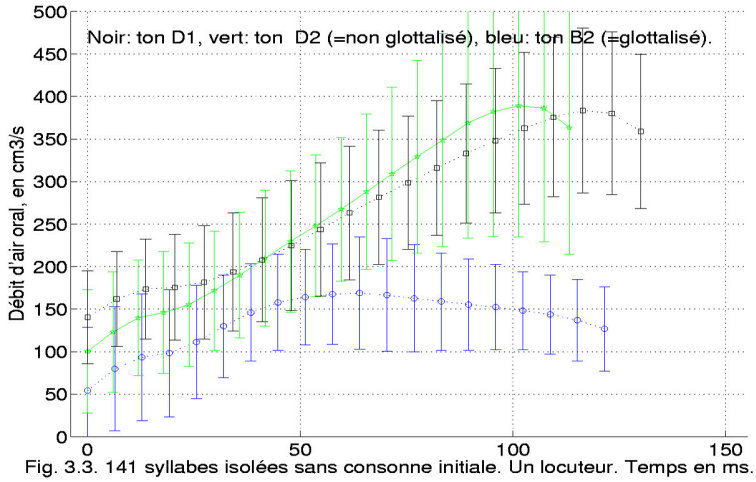


Fig. 3.3. 141 syllabes isolées sans consonne initiale. Un locuteur. Temps en ms.

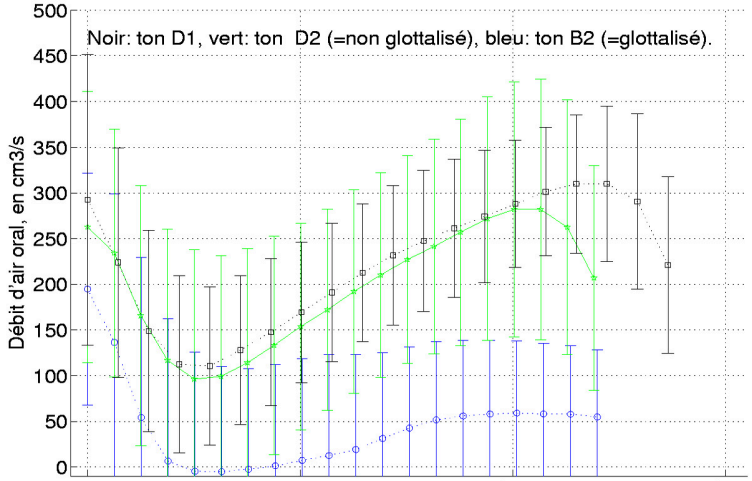


Fig. 3.4. 142 syllabes isolées avec consonne initiale /t/. Un locuteur. Temps en ms

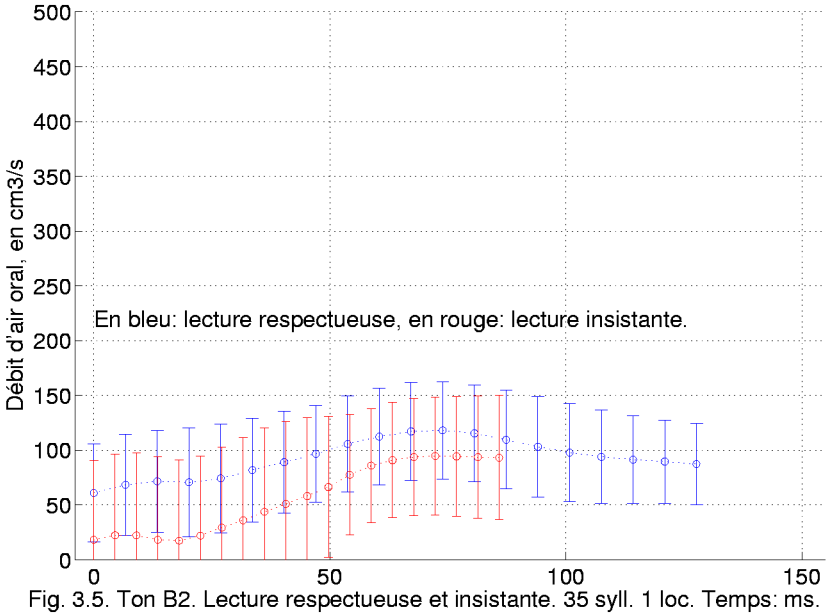


Fig. 3.5. Ton B2. Lecture respectueuse et insistante. 35 syll. 1 loc. Temps: ms.

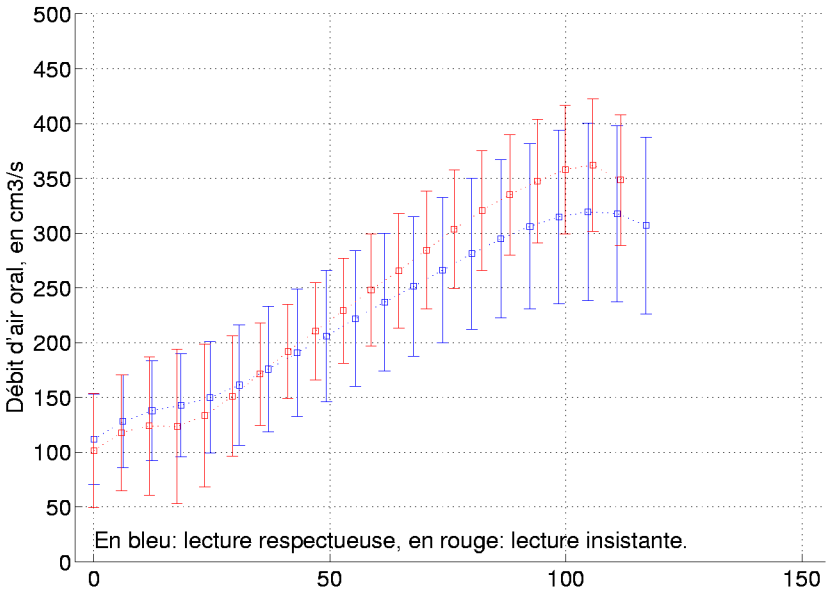


Fig. 3.6. Ton D1. Lecture respectueuse et insistante. 35 syll. 1 loc. Temps: ms.

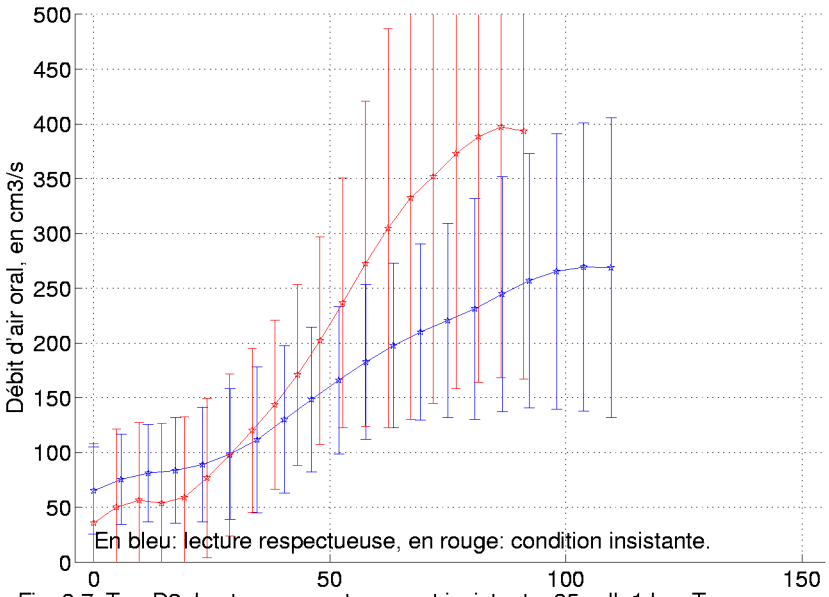


Fig. 3.7. Ton D2. Lecture respectueuse et insistante. 35 syll. 1 loc. Temps: ms.

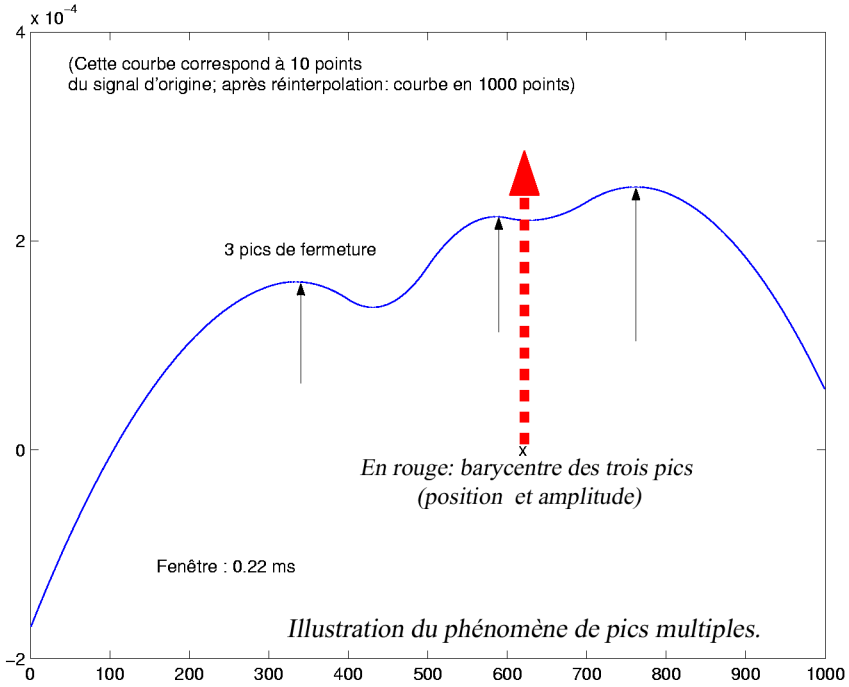


Fig. 4.1. Pic de fermeture en très gros plan, avec indication du barycentre des trois maxima locaux.

Illustration d'une des fenêtres de l'interface de vérification des valeurs de quotient ouvert.

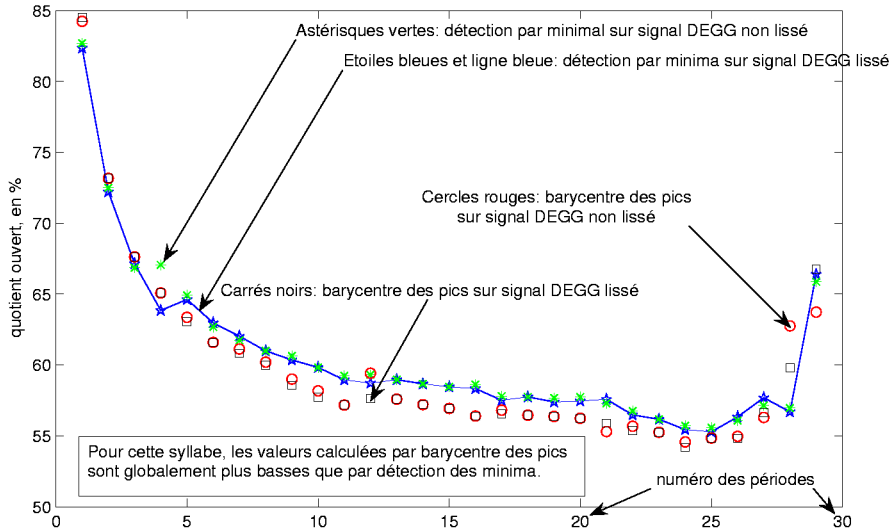


Fig. 4.2. Exemple dans lequel le quotient ouvert calculé par la méthode des barycentres est plus bas que celui obtenu par la détection du minimum local. (Langue naxi, locuteur M9, syllabe 12 du corpus.)

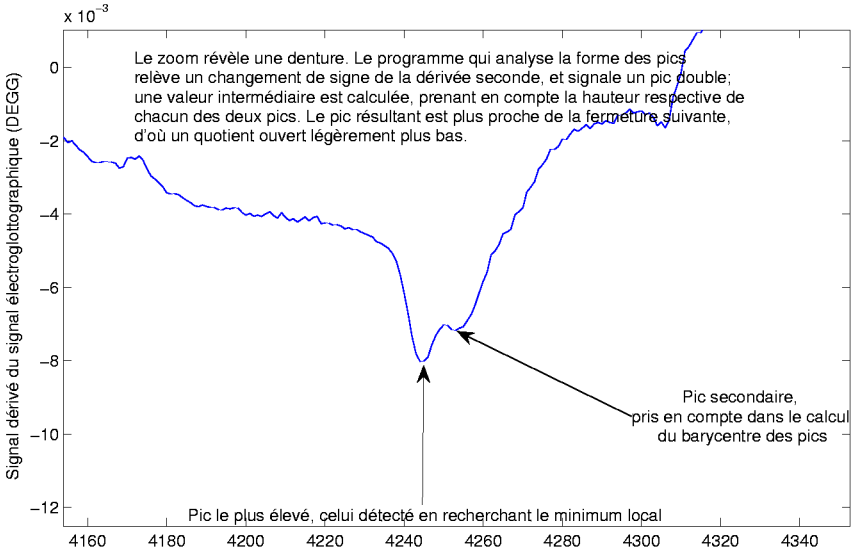


Fig. 4.3. Langue naxi, lecture soignée, syll. 12, loc. M9 : exemple de pic d'ouverture dédoublé. Abscisses: en échantillons.

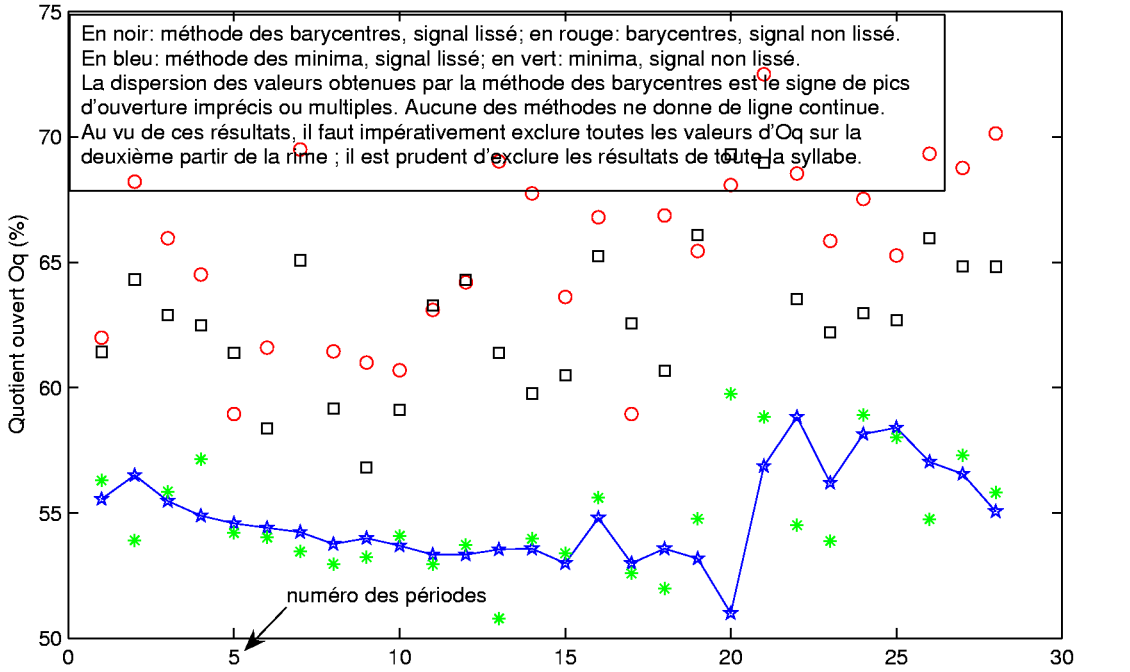


Fig. 4.4. Exemple de résultats de quotient ouvert qui signalent des pics d'ouverture peu marqués ou multiples

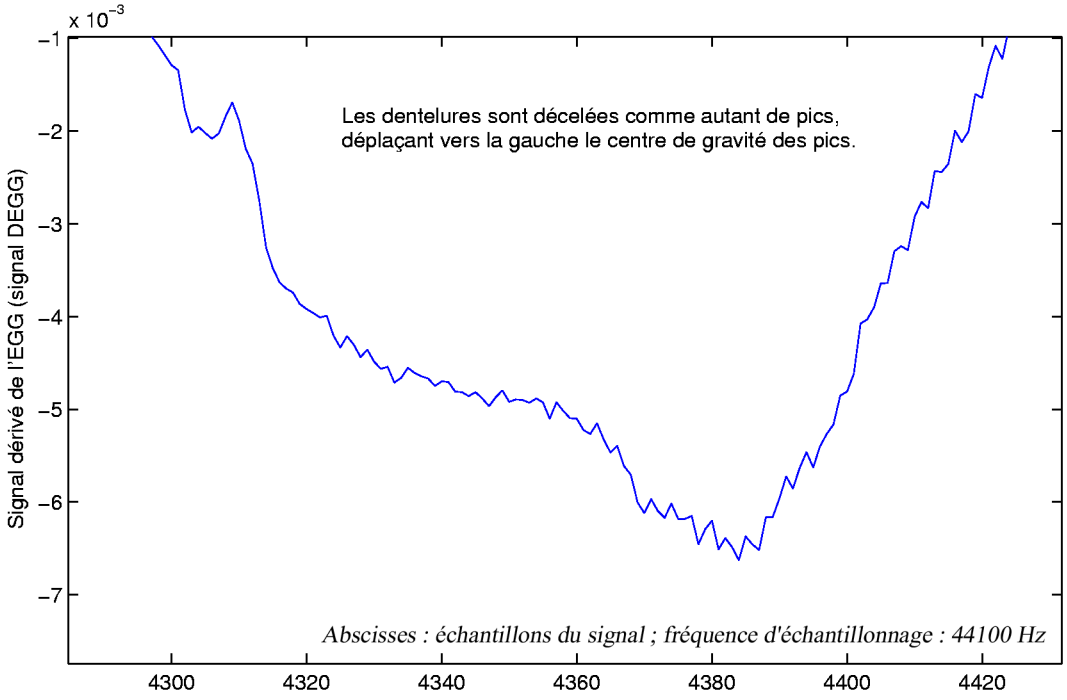


Fig. 4.5. Exemple de pics d'ouverture peu précis, en très gros plan. Syllabe 12, lecture soignée, locuteur naxi M9.

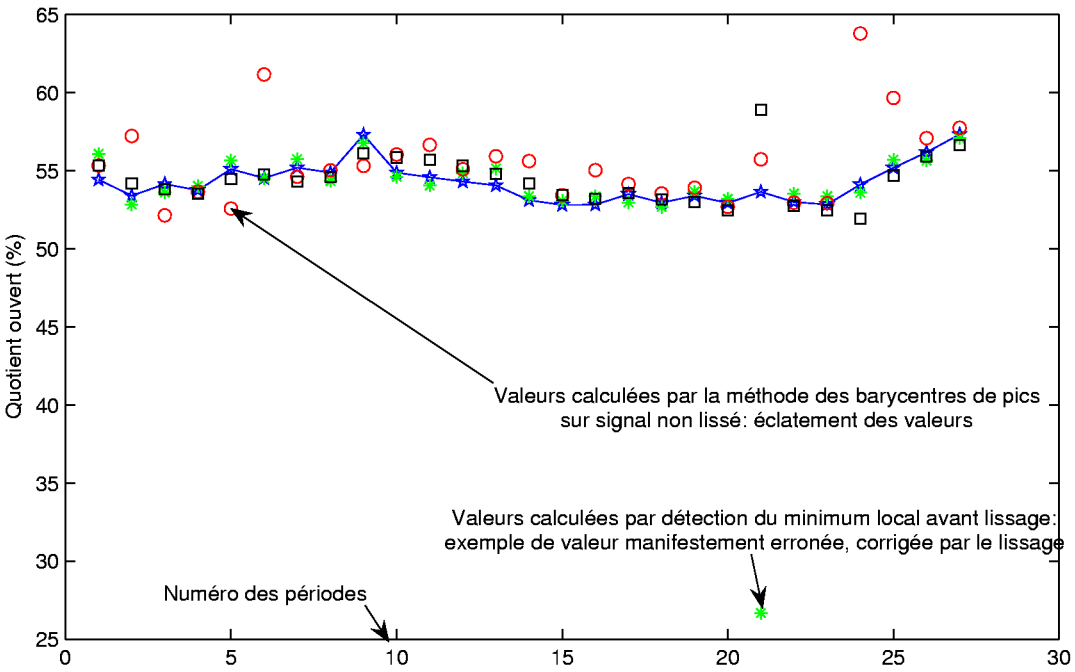


Fig. 4.6. Exemple d'un accord global entre méthodes. Langue: naxi, locuteur M4.

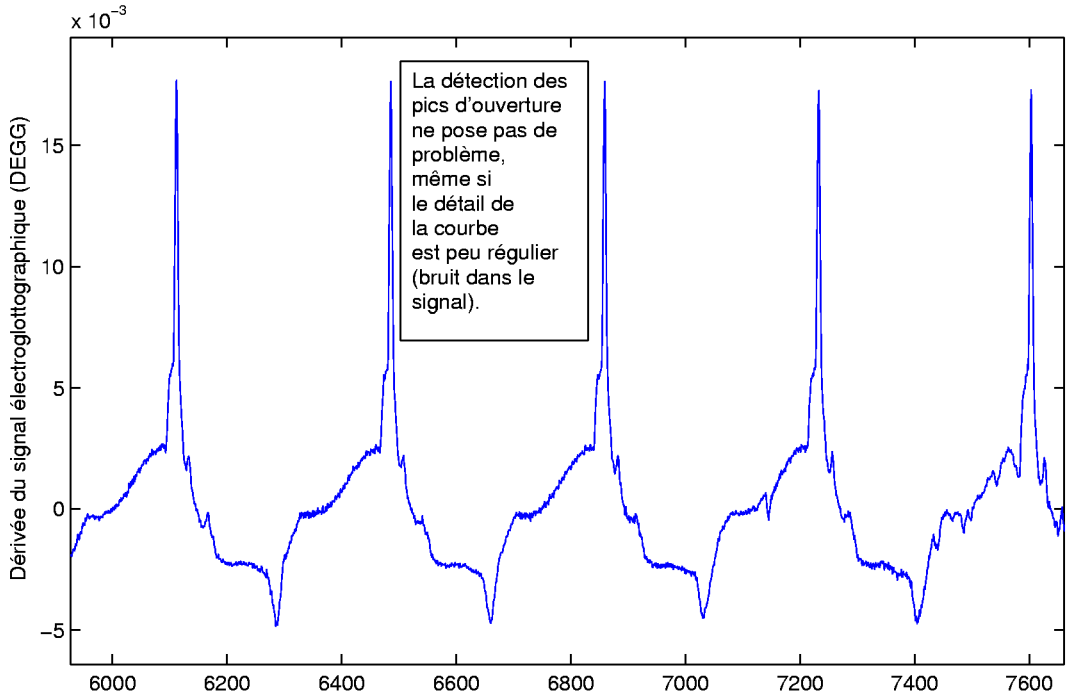


Fig. 4.7. Exemple de pics globalement bien marqués, malgré un signal relativement bruité. (Abscisses: échantillons.)

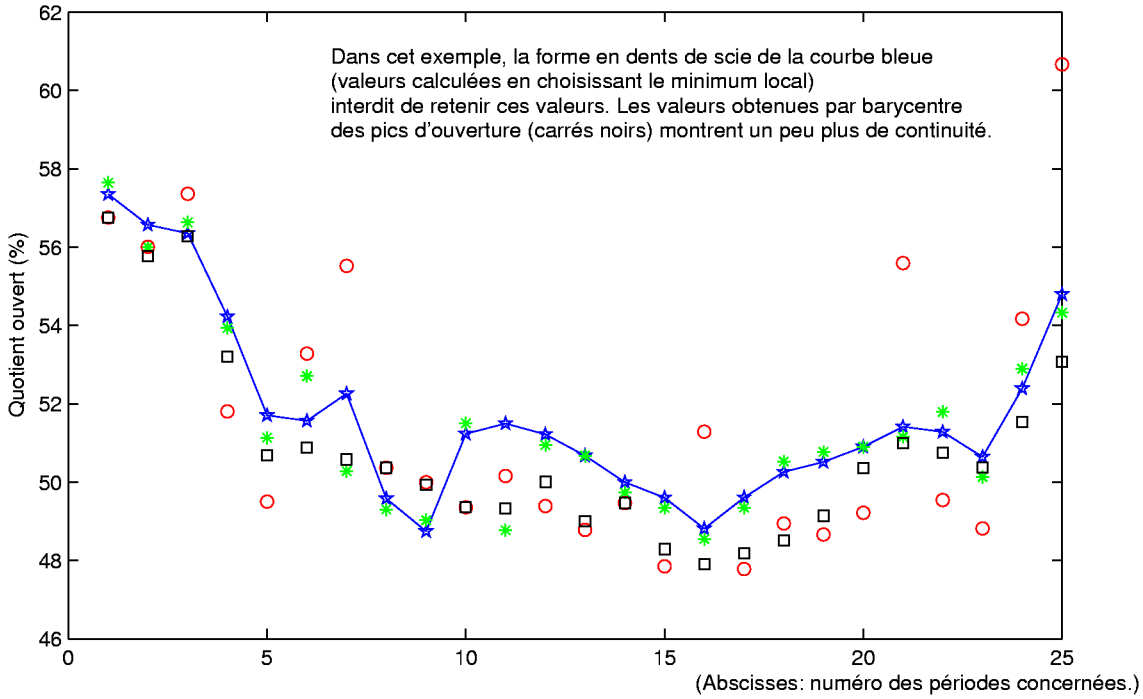


Fig. 4.8. Exemple dans lequel la méthode des barycentres paraît fournir un résultat intéressant.

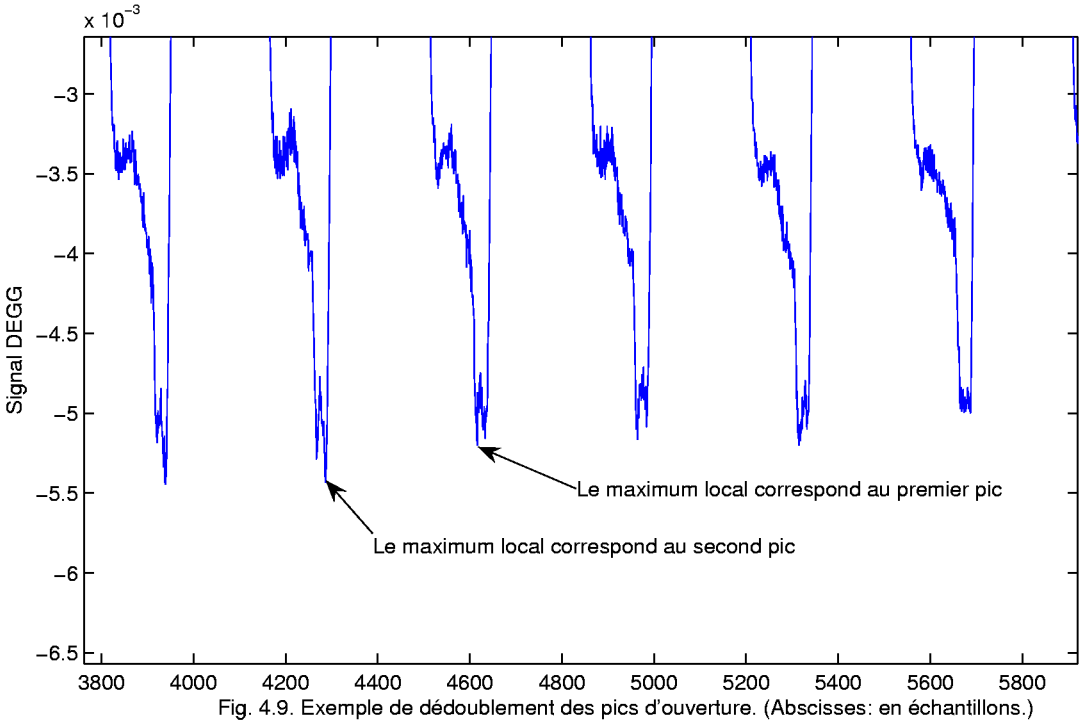


Fig. 4.9. Exemple de dédoublement des pics d'ouverture. (Abscisses: en échantillons.)

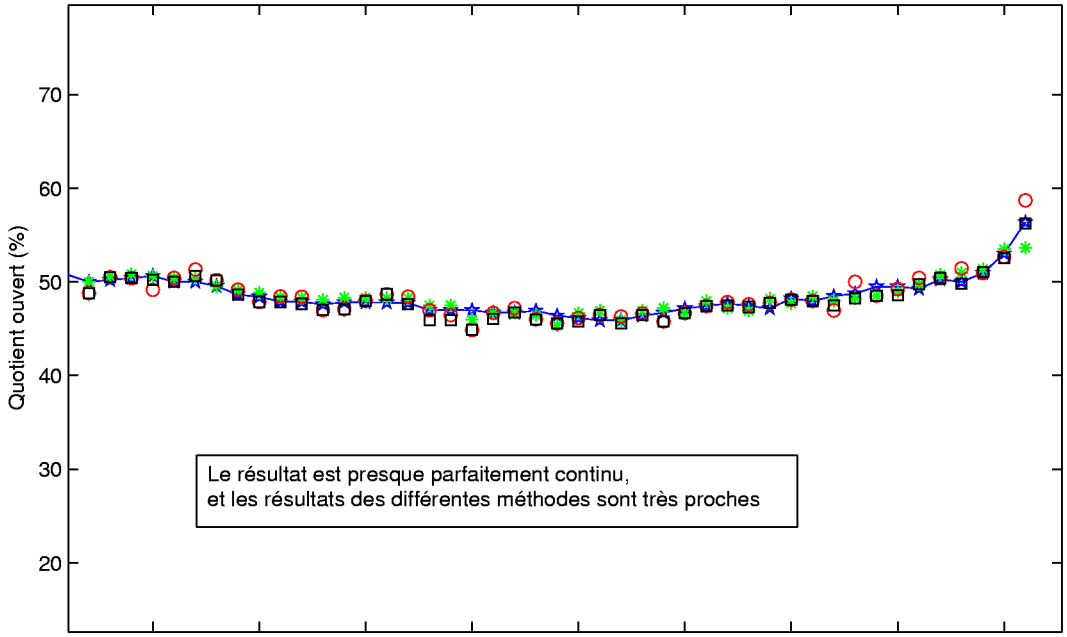


Fig. 4.10. Exemple de résultats considérés comme pleinement satisfaisants

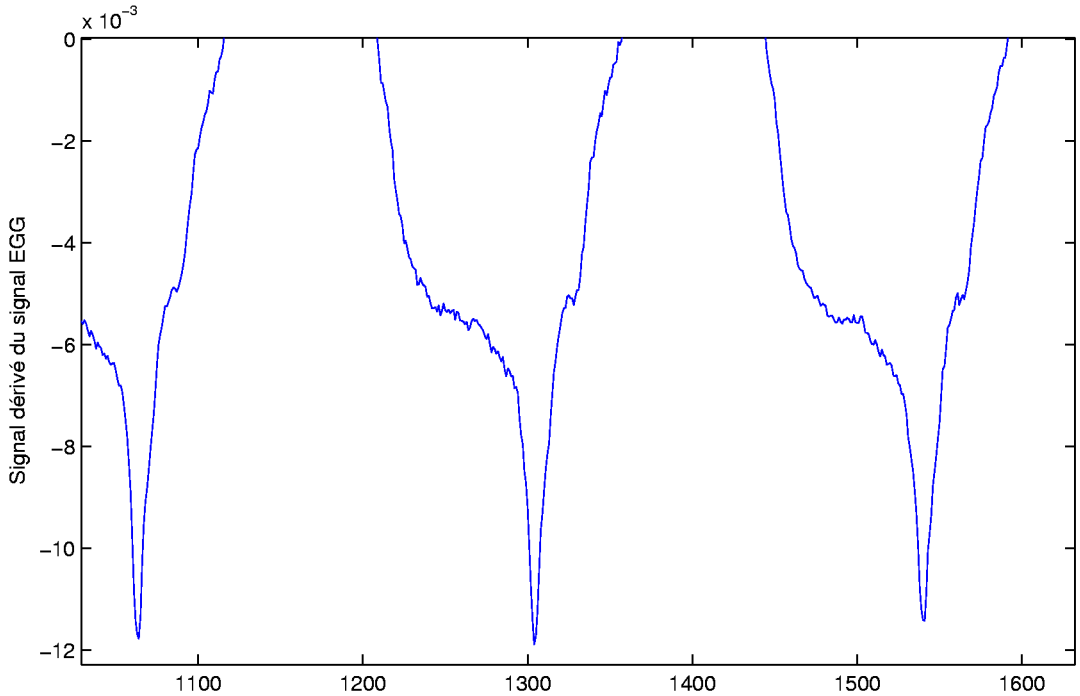


Fig. 4.11. Exemple de pics d'ouverture uniques et bien marqués. (Abscisses: échantillons.)

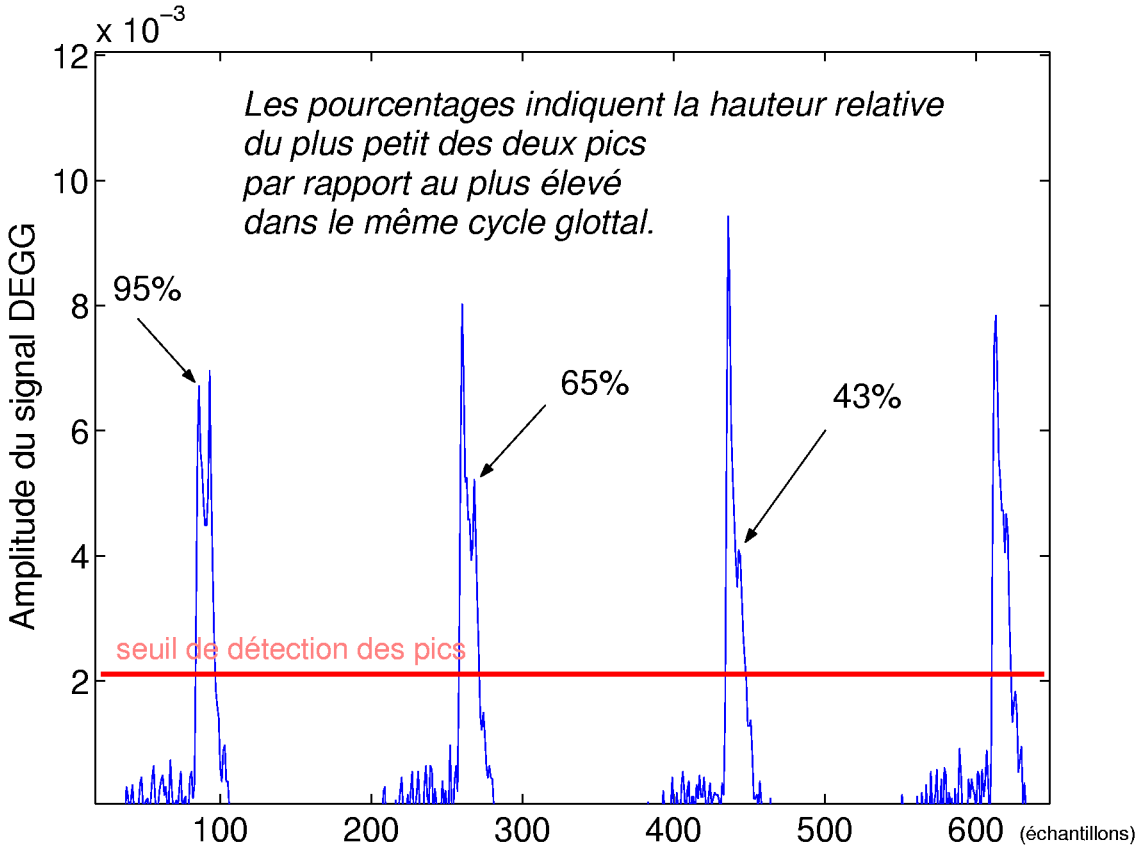


Fig. 4.12. Exemple de doubles pics de fermeture. Locutrice vietnamienne F1. Début de syllabe /i/, ton A1.

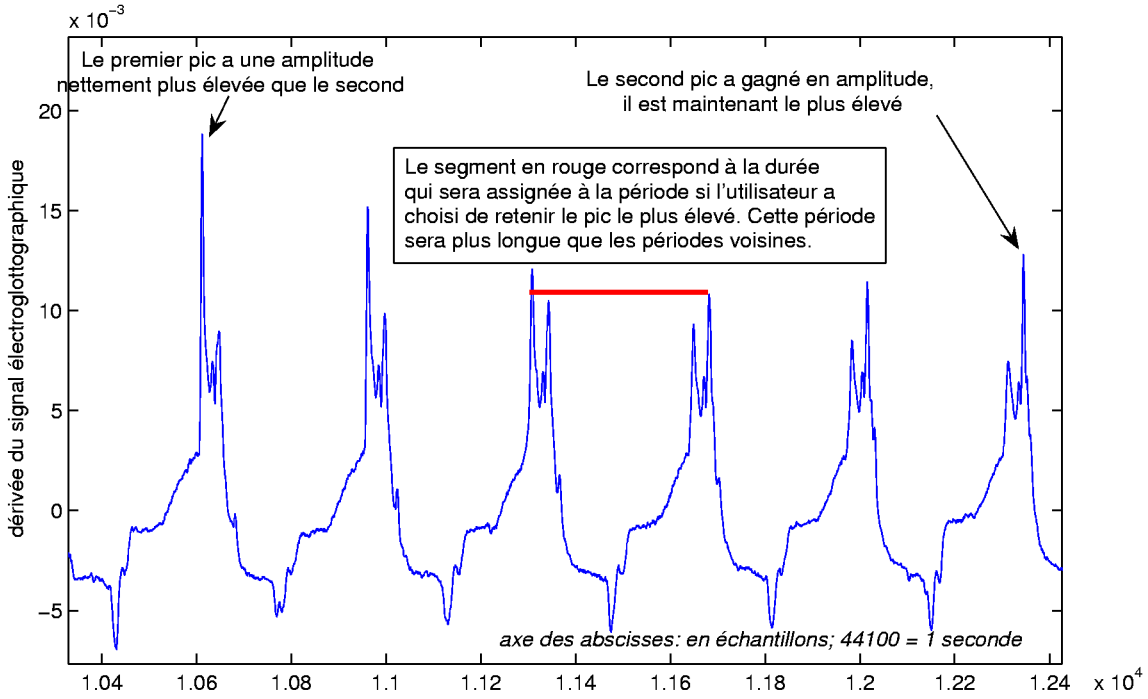


Fig. 4.13. Exemple de changement d'amplitude relative des minima locaux à l'intérieur d'un pic dédoublé, au cours d'une même rime.


/ɑ/
/ə~/
/a/
/v/


/ɤ/


/o/


/e/

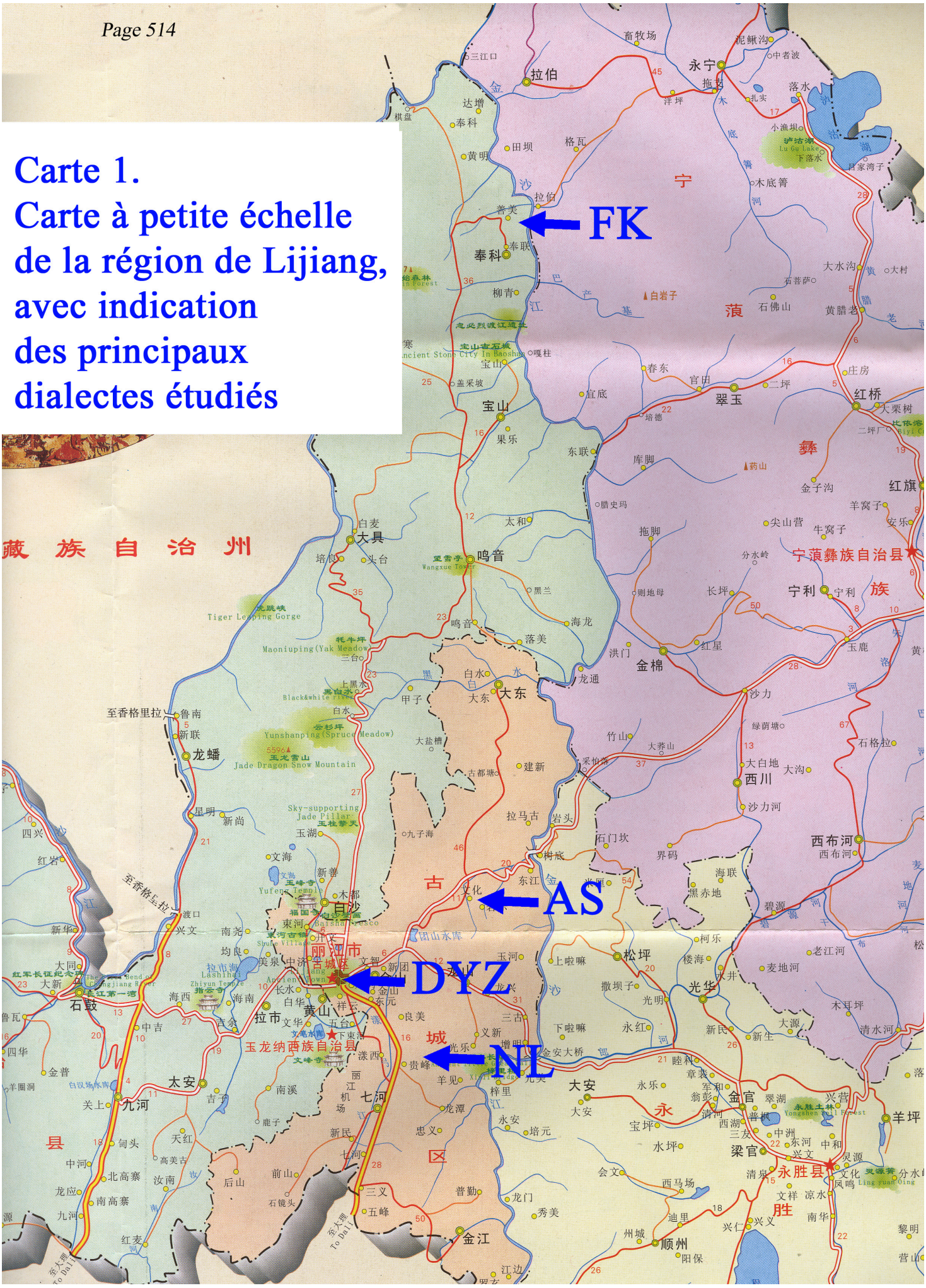

/ɯ/
/u/
/i/
/y/

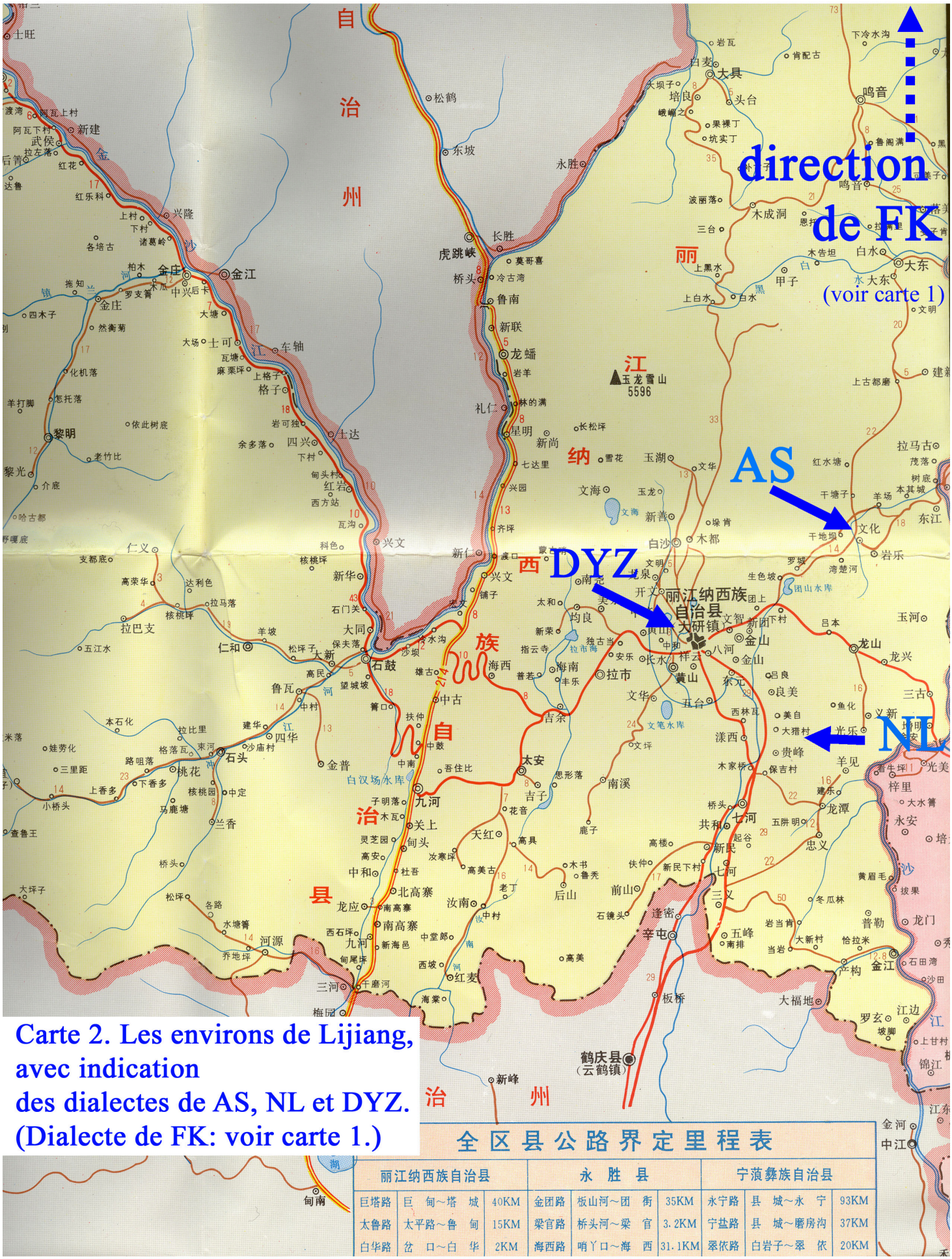

z


z

Cartes

Carte 1.
Carte à petite échelle de la région de Lijiang, avec indication des principaux dialectes étudiés





Carte 2. Les environs de Lijiang, avec indication des dialectes de AS, NL et DYZ. (Dialecte de FK: voir carte 1.)

Index

A

accent d'insistance, 11, 30, 94-97, 101-102, 169, 173-174, 233-234, 402
accent lexical, 1, 2, 8, 12, 39, 131, 202-203, 208, 218, 223-224, 243, 244-245, 248, 397, 413-414, 438, 445-446
accent mélodique, 11, 184, 192-193, 208, 214, 217-218, 221-223, 238-239, 245
accentuation,1, 7-8, 12-13, 99-100, 111, 175-177, 184, 193, 197, 199-201, 203-204, 207, 210, 239-240, 245, 364, 445, 519
analyse autosegmentale, 3, 189, 193, 204, 238
analyse superpositionnelle,187, 192, 198, 199
articulation, 6-7, 31, 52, 68, 72, 83, 108, 112-113, 118, 128, 140, 150, 153, 157-158, 186, 188-189, 191, 194, 215, 240, 317, 328-330, 332, 334, 339, 349, 356, 378, 396, 397, 400-401, 404, 411, 437, 443

C

constriction glottale, 27, 29, 149, 154-156, 167, 210, 214, 216, 237, 426, 518

D

débit d'air, 9, 18, 27, 29, 150, 164, 166, 297, 398, 403-411, 416, 520
déclinaison, 43-47, 53, 54, 55, 61, 63, 70, 75, 77, 83, 84, 85, 87, 111, 190, 205, 230, 238, 241, 246, 520
DECPA, 24, 27, 94, 97-98, 166, 246, 397, 411-412, 414-417, 422, 429, 433, 441, 517, 518, 519
downstep, 54, 93, 205-206, 227-231, 240, 247, 299
durée, 7, 10, 17-19, 21, 23, 46, 64, 70-71, 82, 97, 100-102, 122, 125, 128-130, 132, 148, 154, 156, 159, 160, 163, 165, 168, 179, 184, 188, 194, 203, 302, 380, 387, 398-399, 404, 407-408, 414-415, 418, 428-430, 433-434, 441, 517-518

E

électroglottographie, 17, 19-27, 30, 38, 64-65, 72-75, 81, 83, 90-91, 96-99, 118-120, 125, 154-155, 157-160, 167, 217, 235, 246, 343, 378, 393, 398, 403, 405, 407-408, 411-412, 414-433, 517-519, 521

F

focalisation contrastive, 27, 101, 105, 177-178, 186, 244, 248, 414
focus, 14-15, 27, 77, 89-93, 98, 105, 108, 111, 177-179, 186, 244, 248, 296, 304, 412
formant, 18-19, 31-35, 37, 127-132, 140-141, 185-186, 247, 304, 346-347, 356, 369, 431
frontière (frontière intonative, *intonational boundary, intonational juncture*), 4, 48, 51, 54, 79, 84-85, 116, 119, 187, 190-191, 196, 203, 221, 229, 232, 248, 300-301, 310, 315, 377, 422

G

glottalisation, 9-10, 19, 49, 50, 85, 111, 118, 124, 154, 158, 178, 212, 236-237, 518

H

harmonique, 18, 32-33, 247

I

intensification intonative, 104, 108, 111, 124, 175, 177, 181-182, 207, 244, 414
intensité (intensité acoustique), 7, 10, 15, 17-18, 20, 31, 38-39, 44, 53, 64, 67, 81, 91, 94, 97-99, 101-102, 107, 118, 125-126, 128-133, 145-146, 166-171, 176-177, 180-186, 227, 236, 239, 246-247, 296, 408, 412-415, 438, 441-442, 517-518, 520
intonation pragmatique (composante pragmatique de l'intonation),1, 12, 13, 14, 15, 55, 89, 107, 108, 111, 176, 184, 243, 414-415, 448
intonation syntaxique (composante syntaxique de l'intonation), 1, 2, 11-14, 43-44, 48, 55, 87, 96, 176, 187, 200, 225, 231, 298, 448

L

laryngalisation, 28-29, 68, 94, 148, 154, 156, 210, 216, 237, 411, 518
longueur, 1, 23, 53, 78, 81-82, 125, 126, 132, 133, 139, 144-145, 147, 149, 156, 160, 165, 167-169, 175, 177-179, 185, 202, 208, 239, 351, 393, 396, 398, 401, 403, 408, 434-435, 441

M

mécanisme laryngien, 28-30, 94, 148, 159, 415-416, 422, 517
morphosyntaxe, 4, 13, 76, 80, 91, 203, 207, 210, 225, 240

O

omnisyllabique, 47, 63, 370

P

partage des ressources, 1, 175, 184, 239
pente spectrale, 18-19, 155, 183, 412- 416
perception, 5, 7, 35, 38, 52, 63-66, 68-70, 72, 77, 114, 118, 121, 123, 166, 176, 183, 186, 199, 236, 299, 301, 381, 394, 400, 414, 438, 439, 442
phonème, 4, 9, 39, 47-48, 51, 69, 93, 102, 119, 133, 134, 140, 145, 178-179, 207, 210, 307-308, 310, 315-319, 322, 325-330, 332-336, 343, 350-351, 355, 359, 362, 369, 374, 378-379, 401, 411, 413, 443
phonologisation, 61, 62, 238

Q

qualité de voix, 1, 2, 4, 6, 8-10, 12, 17-18, 22, 83, 106, 117-118, 132, 143, 146-147, 149, 154, 157-159, 164, 167, 175-178, 208, 209-212, 214-215, 217, 225, 235-239, 245, 355-356, 392, 403, 409, 411, 426, 444, 517
quotient ouvert, 22-23, 26-27, 29-30, 64, 67, 81, 83, 85, 93-94, 97-101, 107, 115, 118, 120-121, 124-125, 127, 133, 139, 143-144, 147-148, 155-156, 158-159, 163-169, 177-178, 183, 186, 199, 235, 237, 356, 391-393, 397, 401, 403, 411, 417, 421, 423, 425-429, 432-433, 435, 444, 517-521

R

registre de qualité de voix, 1, 2, 9, 12, 201, 208-210, 245

T

ton de mot, 47, 61, 63, 77, 238, 373, 374
ton ponctuel *(level tone)*, 3, 4, 6, 78, 87, 149, 188, 193, 194, 197, 201, 204-206, 208, 211-213, 218, 221-222, 225-226, 239, 310, 373
transphonologisation, 9, 208-209, 318, 444
typologie, 1-3, 5-7, 11, 14-15, 48, 55, 77-78, 87, 102, 107, 172, 175, 178, 194, 199, 201, 204-206, 218, 221-223, 225-226, 230-233, 235-236, 239-242, 248, 364, 448

V

voix soufflée *(breathy voice)*, 8, 9, 10, 18, 22, 27, 157, 163, 199, 208-210, 215-216, 298, 304, 410-411

Table des figures du volume I

0.1. Courbes moyennes de F_0 des rimes de 192 monosyllabes ; langue : anglais ; 4 locuteurs. _____435

0.2. Le conduit vocal : schéma faisant ressortir la division en trois composantes, sous-glottique, glottique et supra-glottique. (Repris de Catford 1977:18.) _____ 435

1.1 Exemple de coupe spectrale, avec indication de valeurs couramment relevées pour l'estimation de la qualité de voix. (Repris de Wayland et Jongman 2003:190.) _____ 437

1.2 Visualisation de la fermeture glottique par cinématographie ultra-rapide et électroglottographie simultanées. (Repris de Henrich 2001:94.) _____ 438

1.3 Visualisation de l'ouverture glottique par cinématographie ultra-rapide et électroglottographie simultanées. (Repris de Henrich 2001:95.) _____ 439

1.4 Le larynx vu (a) du côté gauche, (b) de l'arrière, (c) de dessus, (d) en section transversale. (Repris de Catford 1977:94.) _____ 440

1.5. Illustration de la méthode d'analyse par seuillage du signal électroglottographique employée par J. Esling. _ 441

1.6. Exemple de signal électroglottographique et de signal dérivé utilisés pour le calcul de F_0 et du quotient ouvert. _____ 441

1.7. Exemple dans lequel les pics de fermeture sont de très faible amplitude. _____ 441

1.8. Signal électroglottographique correspondant aux fermetures de très faible amplitude de la figure 1.7. _ 441

1.9. Exemple de doubles pics de fermeture sur le signal dérivé du signal électroglottographique. _____ 442

1.10. Exemple de changement d'amplitude, du premier au second pic de fermeture, au cours d'une rime. _ 442

1.11 et 1.12. Courbes de F_0 et DECPA sur des phrases en chinois mandarin, illustrant l'effet de changements de mécanisme laryngien sur les valeurs de DECPA. _____ 443

1.13 Le modèle source-filtre de la production de la parole. (Repris de Fant 1960:19.) _____ 443

1.14 Détection des formants pour une voyelle dont le timbre est maintenu, cependant que la fréquence fondamentale monte, de 75 à 175 Hz. (Repris de Fant 1974:38.) _____ 443

1.15. Exemple de l'affichage des résultats de formants dans l'interface de vérification du script <forver>. 443

2.1.a. Courbes moyennes de F_0 et quotient ouvert (normalisées par la durée moyenne) des tons bas et haut dans 10 quadrisyllabes réduplicatifs de schème tonal LHLH, comparées aux courbes moyennes des tons bas et haut sur monosyllabes en phrase-cadre (stylisées d'après 171 items, même locuteur : locuteur naxi M1). _____ 444

2.1.b. Courbes moyennes d'intensité correspondant aux quadrisyllabes réduplicatifs de la figure 2.1a. _____ 444

2.2 à 2.12. Courbes moyennes de F_0 et quotient ouvert des quadrisyllabes réduplicatifs de schème tonal LHLH, locuteurs naxi M4, M5, M7, M8, M9, F2. Position finale d'énoncé : nombres impairs, position non finale : nombres pairs. _____ 445-446

2.13 à 2.28. Courbes moyennes de F_0 et quotient ouvert d'énoncés dont toutes les syllabes portent le même ton (H, M, L ou LH). Nombres impairs : courbes moyennées, nombres pairs : courbes stylisées en deux points. ___ 447-454

2.29. Courbes de F_0 et quotient ouvert de l'expression « un lac ». Locuteur naxi M1. _____ 455

2.30. Courbes de DECPA et d'intensité de l'expression « un lac ». Locuteur naxi M1. _____ 455

2.31. Courbes de F_0 et quotient ouvert de la phrase 23 (« Tu vas jusqu'à Pékin ? »). Locuteur naxi M1. ___ 456

2.32. Courbes de DECPA et d'intensité de la phrase 23 (« Tu vas jusqu'à Pékin ? »). Locuteur naxi M1. _ 456

2.33. Courbes de F_0, quotient ouvert et DECPA de la phrase 24 (« C'est à Pékin que je vais »). Locuteur naxi M1. _____ 457

2.34. Courbes de F_0 et quotient ouvert de la phrase 17' (« Non, c'est par... »). Locuteur naxi M1. _____ 457

2.35. Courbes de DECPA et d'intensité de la phrase 17' (« Non, c'est par... »). Locuteur naxi M1. _____ 457

3.1 à 3.12. Courbes moyennes de F_0 et quotient ouvert des mots-cible en anglais, permettant une comparaison entre conditions de lecture. Monosyllabes (figures à numéros impairs) et disyllabes (figures à numéros pairs), locuteurs anglais M2 à M7. _____ 458-463

3.13a-b à 3.27a-b. Estimation des trois premiers formants des voyelles des monosyllabes en anglais. Résultats bruts (figures 3.13a, 3.14a...) et après suppression de valeurs aberrantes suite à vérification visuelle (figures 3.13b, 3.14b...), locuteurs anglais M3 à M7. _____ 464-468

3.28 à 3.30. Courbes moyennes d'intensité acoustique globale des mots-cible en anglais, permettant une comparaison entre conditions de lecture. Monosyllabes, voyelle /ɑː/ (3.28), /iː/ (3.29), /uː/ (3.30). Locuteur anglais M4. _____ 469

3.31 à 3.45. Courbes moyennes de F_0 et quotient ouvert des mots-cible en naxi, permettant une comparaison entre conditions de lecture. Locuteurs naxi M5, M7, M8, M9, et F2. _____ 470-474

3.46 à 3.55. Fréquences formantiques des voyelles naxi (dans un espace F1-F2 sur les figures à numéro impair, F2-F3 sur les figures à numéro pair), contrastant la condition de lecture soignée (en bleu) et la condition de lecture insistante (en rouge) ainsi que les trois tons : voyelles au ton H représentées par un caractère de grande taille, au ton M par un caractère de taille moyenne, au ton L par un caractère de petite taille. Locuteurs naxi M5, M7, M8, M9, et F2. _____ 475-479

3.56. Exemple de représentation de voyelles en trois dimensions (F1, F2, F3). Locuteur naxi M3. _____ 480

3.57. Signal audio et électroglottographique (avec sa dérivée) de la fin d'une syllabe au ton B2, illustrant la constriction glottale. _____ 480

3.58. Signal audio et électroglottographique (avec sa dérivée) de la fin d'une syllabe au ton C1, fournissant un exemple de laryngalisation. _____ 480

3.59. Signal audio et électroglottographique (avec sa dérivée) de la fin d'une syllabe au ton D2, illustrant l'absence de glottalisation. _____ 480

3.60. Même exemple que sur la fig. 3.59 (fin de syllabe /ap/, locuteur M1), avec repère correspondant approximativement à l'occlusion labiale. _____ 481

3.61. Signal audio et électroglottographique des dernières périodes de la syllabe /i/, ton A2, avant pause, illustrant la fin progressive des oscillations sur le signal électroglottographique. Locutrice vietnamienne F1. _____ 481

3.62. Spectrogrammes de syllabes aux tons B2 et D2. _____ 481

3.63. Signal audio et électroglottographique de la fin de la syllabe /am/ au ton B2, locuteur vietnamien M1. 482

3.64. Signal audio et électroglottographique fournissant l'exemple d'un contact irrégulier des plis vocaux au ton B2. Locuteur vietnamien M1. _____ 482

3.65. Signal audio et électroglottographique fournissant l'exemple d'un train de cycles glottiques de faible amplitude au ton B2. Locuteur vietnamien M1. _____ 482

3.66. Signal électroglottographique illustrant l'impossibilité d'estimer la durée de la phase fermée après la dernière fermeture en fin de voisement. _____ 482

3.67. Ton B2 par le locuteur vietnamien M1. 3.67a. Courbes brutes de F_0 et quotient ouvert de 42 items, lecture soignée. 3.67b. Courbes brutes de F_0 et quotient ouvert de 42 items, lecture insistante. 3.67c. Courbes moyennes de F_0 et quotient ouvert. _____ 483

3.68. Ton D2 par le locuteur vietnamien M1. 3.68a. Courbes brutes de F_0 et quotient ouvert de 42 items, lecture soignée. 3.68b. Courbes brutes de F_0 et quotient ouvert de 42 items, lecture insistante. 3.68c. Courbes moyennes de F_0 et quotient ouvert. _____ 484

3.69. Courbes moyennes de F_0 et quotient ouvert au ton D1, locuteur vietnamien M1. _____ 485

3.70. Ton B2 par le locuteur vietnamien M2. 3.70a. Courbes brutes de F_0 et quotient ouvert de 42 items, lecture soignée. 3.70b. Courbes brutes de F_0 de 42 items, lecture insistante. 3.70c. Courbes brutes de quotient ouvert de 42 items, lecture insistante. 3.70d. Courbes moyennes de F_0 et quotient ouvert. _____ 485

3.71. Tons D1 et D2 par le locuteur vietnamien M2. 3.71a. Courbes brutes de F_0 et quotient ouvert de 42 items au ton D2, locuteur M2, lecture soignée. 3.71b. Courbes brutes de F_0 et quotient ouvert de 42 items au ton D2, locuteur M2, lecture insistante. 3.71c. Courbes moyennes de F_0 et quotient ouvert du ton D2. 3.71d. Courbes moyennes de F_0 et quotient ouvert du ton D1. _____ 486

3.72. Courbes moyennes de F_0 et quotient ouvert du ton B2. Locuteur vietnamien M3. 3.72a. Lecture soignée, et lecture insistante. 3.72b. Lecture soignée, et lecture respectueuse (condition-contrôle). _____ 486

3.73. Courbes de F_0 et quotient ouvert du ton D2. Locuteur vietnamien M3. 3.73a. Courbes moyennes en lecture soignée et lecture insistante. 3.73b. Courbes moyennes en lecture soignée, et lecture respectueuse (condition-contrôle). 3.73c. Courbes brutes, lecture respectueuse (condition-contrôle). _____ 487

3.74. Courbes moyennes de F_0 et quotient ouvert du ton D1. Locuteur vietnamien M3. 3.74a. Lecture soignée, et lecture insistante. 3.74b. Lecture soignée, et lecture respectueuse (condition-contrôle). _____ 487

3.75. Courbes moyennes de F_0 et quotient ouvert du ton B2. Locutrice vietnamienne F1. _____ 488

3.76. Courbes moyennes de F_0 et quotient ouvert du ton D2. Locutrice vietnamienne F1. _____ 488

3.77. Courbes moyennes de F_0 et quotient ouvert du ton D1. Locutrice vietnamienne F1. _____ 488

3.78. Courbes moyennes de DECPA (moyenne de 84 rimes), ton vietnamien D2. Locuteur vietnamien M1. ____ 489

3.79. Courbes moyennes de DECPA (moyenne de 84 rimes), ton vietnamien B2. Locuteur vietnamien M1. ____ 489

4.1. Schéma des tons du vietnamien de Hanoi, fondé sur les données du locuteur M1. _____ 490

4.2. Place de l'accentuation dans la prosodie chez M. Hoa (haut) et J. Vaissière (bas). _____ 490

Table des photographies et des cartes

Photographies de locuteurs naxi portant les électrodes de l'électroglottographe _____ 436

Photographies du visage du locuteur naxi M4 de face pendant la tenue de voyelles (p. 510 : /ɑ/, /ə˞/, /a/, /ɻ/ ; p. 511 : /ɤ/, /o/, /e/ ; p. 512 : /ɯ/, /u/, /i/, /y/ ; p. 513 : syllabes [z] et [z̩], allophones de /ɯ/) _____ 510-513

Carte 1. Carte à petite échelle de la région de Lijiang, avec indication des principaux dialectes étudiés. ___ 514

Carte 2. Les environs de Lijiang, avec indication des dialectes de AS, NL et DYZ. _____ 515

Table des graphiques

2.1 et 2.2. Résultats du test d'identification du ton des syllabes isolées extraites des séquences réduplicatives A1 B1 A2 B2. 40 stimuli (signal électroglottographique), 15 auditeurs ; données produites par le locuteur M1 (graphique 2.1) et par M2 (graphique 2.2). _____ 491

2.3. Identification du ton des syllabes extraites des séquences réduplicatives A1 B1 A2 B2 lorsqu'elles figurent en première position d'un stimulus disyllabique. 140 stimuli, 14 auditeurs. _____ 491

2.4 et 2.5. Identifications du ton des syllabes extraites des séquences réduplicuées A1 B1 A2 B2 en fonction de la position dans le composé disyllabique utilisé comme stimulus. 140 stimuli, 14 auditeurs ; données produites par le locuteur M1 (graphique 2.1) et par M2 (graphique 2.2). _____ 491

2.6. Approximation en tons musicaux de la déclinaison à l'échelle de l'énoncé dans les énoncés naxi dont toutes les syllabes portent un même ton lexical. Chaque point correspond à la moyenne de plusieurs énoncés (voir détails sur les figures 2.13 à 2.27). Chaque couleur correspond à un locuteur. _____ 491

3.1 à 3.9. Différences moyennes de fréquence fondamentale et quotient ouvert entre conditions de lecture, en fonction du locuteur et du type de syllabe. Graphiques 3.1 à 3.3 : données anglaises, 3.4 à 3.6 : données naxi, 3.7 à 3.9 : données vietnamiennes. _____ 3.1, 3.2, 3.4, 3.5, 3.7, 3.8 : p. 492 ; 3.3, 3.6, 3.9 : p. 493

3.10. Bilan simplifié des différences entre conditions de lecture dans les trois langues. _____ 493

3.11. Rapport entre variation de F_0 et variation d'intensité acoustique globale : d'une condition de lecture à l'autre, dans les trois langues (en bleu), et entre tons lexicaux M et H en naxi (en rouge). _____ 493

Table des figures des Annexes

1.1. Les phonèmes vocaliques simples du naxi de AS. Trois degrés d'aperture, deux degrés d'antériorité, une opposition d'arrondissement ; un /ɥ/ syllabique et une voyelle rhotique. _____ 494

1.2. Estimation des fréquences des deux premiers formants des principales voyelles naxi. Locuteur M3. ___ 494

2.1. Courbes moyennes de F_0 et quotient ouvert de 317 disyllabes, locuteur anglais M1, lecture soignée et insistante. _____ 495

2.2. Courbes brutes de F_0 de 24 monosyllabes, distinguant entre trois catégories d'initiales ; locuteur anglais M4, lecture soignée. _____ 496

2.3 à 2.10. Courbes moyennes de F_0 distinguant entre trois catégories d'initiales ; locuteurs anglais M4 à M7, énonciation soignée et énonciation insistante. 24 syllabes par figure. _____ 496-499

2.11 à 2.14. Courbes brutes et moyennes de F_0 distinguant entre trois catégories de voyelles ; locuteurs anglais M4 à M7, énonciation soignée et énonciation insistante. 24 syllabes par figure. _____ 500-501

3.1 à 3.4. Mesures de débit d'air oral sur syllabes vietnamiennes aux tons D1, D2, B2. 3.1 : énonciation *respectueuse* ; 3.2 : énonciation *insistante* ; 3.3 : lecture à l'isolée ; 3.4 : lecture à l'isolée, la rime étant précédée d'une consonne /t/. Un locuteur. _____ 502

3.5. Mesures de débit d'air oral sur syllabes vietnamiennes au ton B2, contrastant lecture respectueuse et insistante. Un locuteur. _____ 503

3.6. Mesures de débit d'air oral sur syllabes vietnamiennes au ton D1, contrastant lecture respectueuse et insistante. Un locuteur. _____ 503

3.7. Mesures de débit d'air oral sur syllabes vietnamiennes au ton D2, contrastant lecture respectueuse et insistante. Un locuteur. _____ 503

4.1. Illustration du phénomène de pics multiples. Pic de fermeture en très gros plan, avec indication du barycentre des trois maxima locaux. _____ 504

4.2. Illustration d'une des fenêtres de l'interface de vérification des valeurs de quotient ouvert. Exemple dans lequel le quotient ouvert calculé par la méthode des barycentres est plus bas que celui obtenu par la détection du minimum local. (Langue naxi, locuteur M9, syllabe 12 du corpus.) _____ 504

4.3. Exemple commenté de pic d'ouverture dédoublé. _____ 504

4.4. Exemple de résultats de quotient ouvert qui signalent des pics peu marqués ou multiples. _____ 505

4.5. Exemples de pics d'ouverture peu précis, en très gros plan. Langue naxi, locuteur M9, syllabe 12 du corpus, énonciation *soignée*. _____ 505

4.6. Exemple dans lequel l'accord entre méthodes de calcul du quotient ouvert est bon dans l'ensemble. Langue : naxi, locuteur M4. _____ 506

4.7. Exemple de pics relativement bien marqués, malgré un signal relativement bruité. _____ 506

4.8. Exemple dans lequel la méthode des barycentres paraît fournir un résultat intéressant pour une estimation du quotient ouvert. _____ 507

4.9. Exemple de dédoublement des pics d'ouverture. _____ 507

4.10. Exemple de résultats considérés comme pleinement satisfaisants. _____ 508

4.11. Exemple de pics d'ouverture uniques et bien marqués. _____ 508

4.12. Exemple de doubles pics de fermeture, avec indication sur les résultats de la procédure de détection des pics multiples _____ 509

4.13. Exemple de changement d'amplitude relative des minima locaux à l'intérieur d'un pic dédoublé, au cours d'une même rime _____ 509

Table des matières détaillée

Remerciements

Table des matières brève

Conventions concernant les citations en langues étrangères

Mise en page et conventions typographiques

Introduction 1

1. Problématique de recherche : l'hypothèse d'un *partage des ressources* entre l'accentuation et l'intonation 1

2. Enjeu au plan de la modélisation prosodique et de la typologie prosodique 2

3. Présentation du cadre global de recherche 4

3.1. Complémentarité entre phonétique et phonologie 5

3.2. Variété des langues et synthèse typologique 6

3.3. Multiplicité des paramètres prosodiques, et utilité des explorations physiologiques 6

3.3.1. Rôle synchronique et diachronique de la qualité de voix dans les systèmes tonals 8

3.3.2. Association fréquente d'une certaine qualité de voix à des phénomènes intonatifs 9

4. Notions centrales de la modélisation prosodique 10

4.1. Cadre britannique 10

4.2. Cadre de la présente étude 11

4.2.1. Intonation syntaxique (= composante pragmatique de l'intonation) 12

4.2.2. Intonation pragmatique (= composante pragmatique de l'intonation) 13

Bilan de l'Introduction : Questions centrales 14

Chapitre I. Méthode expérimentale : enjeux et outils 17

1. Intérêt d'une mesure de la qualité de voix (et de la fréquence fondamentale) par électroglottographie 17

1.1. Difficulté des mesures de la qualité de voix (et secondairement de la fréquence fondamentale) à partir du signal audio 17

1.1.1. Difficulté de l'évaluation de la qualité de voix à partir du signal audio 17

1.1.2. Remarques sur la mesure de F_0 à partir du signal audio 19

1.2. Présentation de l'instrument : principes, histoire, et panorama des méthodes d'analyse du signal électroglottographique 19

1.2.1. Principes et histoire 19

1.2.2. Les méthodes de calcul de coefficients par seuillage du signal électroglottographique présentent une part d'incertitude 20

1.2.3. Quotient ouvert et fréquence fondamentale sont calculables à partir du signal dérivé du signal électroglottographique 22

1.2.3.1. Limites de la mesure, du fait des doubles pics de fermeture et d'ouverture 23

1.2.3.2. Limites de la méthode combinée de calcul du quotient ouvert, dite *de Howard* 25

1.2.3.3. Interprétation du quotient ouvert 26

1.2.4. Réalisation pratique des mesures 29

2. Remarques sur les fréquences formantiques et l'intensité 29
2.1. Les fréquences formantiques et leur mesure 29
2.1.1. Rappels théoriques 29
2.1.2. Difficulté et limites de l'estimation des fréquences de résonance du conduit vocal à partir du signal audio 30
2.1.3. Choix d'une procédure pour l'estimation des fréquences formantiques 33
2.1.3.1. Principes de fonctionnement des outils de détection des formants proposés par le logiciel PRAAT 33
2.1.3.2. Choix d'effectuer une vérification visuelle des résultats 34
2.1.3.3. Emploi d'un second logiciel, pour obtenir un point de comparaison 35
2.2. Mesures d'intensité acoustique globale 36
2.2.1. Variabilité en fonction de la position du locuteur par rapport au micro 36
2.2.2. L'interprétation de la mesure d'intensité 37

Chapitre II. Aperçu d'ensemble de l'intonation du naxi 41

Premier volet : Recherches sur l'*intonation syntaxique* de la langue naxi 41

1. Déclinaison et frontières : réflexion sur les notions, et observations préliminaires 42
1.1. Observations au niveau du mot 44
1.2. Observations au niveau de groupements intermédiaires entre le mot et l'énoncé 45
1.2.1. Evaluation de l'abaissement de fréquence fondamentale en fin d'unité intonative dans un récit 46
1.2.2. Deux exemples fournissant une illustration complémentaire de l'effet des frontières sur la réalisation du ton de la syllabe finale d'unité intonative 48
1.3. Observations au niveau de l'énoncé 49
1.4. Observations au niveau d'unités supérieures à l'énoncé 50

2. Réduplication et prosodie : données synchroniques et hypothèse diachronique concernant les formes réduplicatives en naxi 52
2.1. Exposition des principaux faits, et premières hypothèses sur leur origine 53
2.1.1. Bref panorama sémantique des phénomènes de réduplication en naxi 53
2.1.2. Forme phonologique de la réduplication 54
2.1.3. Aucune explication phonologique évidente ne permet de rendre compte des schémas observés 56
2.2. L'hypothèse retenue ici au sujet de l'origine du changement : phonologisation de l'effet du marquage des frontières, à l'occasion d'une transition inaboutie vers un système prosodique à ton de mot 58
2.2.1. Le constat phonétique d'un jeu sur la ligne de déclinaison 58
2.2.2. Hypothèse diachronique 59
2.3. Volet expérimental 60
2.3.1. Méthode : étude combinée de données de production et de perception 61
2.3.2. Les données de production montrent l'ampleur de la variation intonative dans la réalisation des tons (« variation allotonique ») 63
2.3.3. Les tests de perception confirment que la différence entre allotones peut, dans les conditions de l'expérience, être perçue comme une différence catégorielle 65
2.3.3.1. Première partie du test : Syllabes isolées 68
2.3.3.2. Deuxième partie du test : Paires de syllabes 69
2.3.3.3. Troisième partie du test : Quadruplets de syllabes 72

3. Approche de la ligne de déclinaison et des frontières par l'emploi d'énoncés dont toutes les syllabes portent le même ton 74
3.1. Méthode : matériau linguistique et enregistrements 75
3.1.1. Matériau linguistique 75
3.1.2. Mesures de F_0 et quotient ouvert réalisées 77

3.1.3. Représentation et analyse des résultats : résumer les courbes à deux points par syllabes n'apparaît pas adéquat, du fait de la présence de phénomènes d'allongement doublés d'un abaissement final 78
3.2. Résultats : estimation de la déclinaison, et de l'effet des frontières 80
3.3. Premières remarques en vue d'une comparaison de la déclinaison entre langues 82
Bilan général au sujet de la composante syntaxique de l'intonation en naxi 83

Tableau de l'intonation du naxi, deuxième volet : Recherches sur l'intonation pragmatique de la langue naxi *85*

1. Moyens grammaticaux de structuration de l'information 86
1.1. Mise en valeur et mise en retrait 86
1.2. Le rôle de l'ordre des mots 88
1.3. Le rôle des particules grammaticales 88

2. Réalisation d'une intensification par des modifications intonatives locales 89
2.1. Premières observations sur la malléabilité des tons 89
2.2. Le conflit entre ton lexical et mise en valeur intonative 90
2.2.1. Première stratégie de mise en valeur : l'intonation l'emporte sur le ton 91
2.2.2. Deuxième stratégie de mise en valeur (observé seulement sur des voyelles d'arrière) : serrage pharyngien 92
2.2.3. Troisième statégie de mise en valeur : effort supplémentaire qui n'est pas signalé uniquement par un changement de F_0. Choix de la placer au centre de l'expérience principale (rapportée au chapitre III) 92
2.3. Réflexions sur la notion d'*accent d'insistance* 94
2.3.1. Caractérisation de l'accent d'insistance en termes de quotient ouvert : il n'est pas accompagné par une adduction forte des plis vocaux 94
2.3.2. L'accent d'insistance, universel de l'intonation ? 95
2.4. L'ajout d'un ton H intensif, processus catégoriel 98
2.4.1. Description générale du processus 99
2.4.2. L'ajout d'un ton H intensif est distinct de la focalisation contrastive (absente en naxi) 100
2.4.3. L'ajout de ton H intensif est distinct de l'accent d'insistance 101
2.4.4. Perspective typologique au sujet du ton H intensif 102
2.5. Bilan des expériences sur la composante pragmatique de l'intonation en naxi 103

Chapitre III. Etude comparée des modifications locales induites par l'intonation pragmatique en naxi, vietnamien et anglais *107*

Commentaires de méthode sur la tâche de production 107
Utilité d'une vérification visuelle des mesures 110
Une limite : l'absence de vérification perceptive 110

1. Données anglaises 111
1.1. Expérience-pilote 111
1.1.1. Mots retenus pour la première étape (étude-pilote) 111
1.1.2. Phrases-cadres 112
1.2. Mots choisis pour l'expérience principale 113
1.3. Attentes au sujet de la réalisation des mots-cible 115
1.4. Locuteurs et déroulement de l'enregistrement 115
1.5. Bilan des enregistrements 116
1.6. Représentation visuelle des données 116

1.7. Tableaux de résultats pour les locuteurs principaux (M3 à M7) 116
1.7.1. F_0 moyenne : la syllabe accentuée est plus fortement modifiée 116
1.7.2. Evolution de F_0 : l'allure de la courbe change dans la moitié des cas 118
1.7.3. Le quotient ouvert est plus bas en lecture insistante 119
1.7.4. Longueur de la rime : variété des stratégies 120
1.7.5. Peu de variation des formants, augmentation d'intensité acoustique globale 122
1.7.6. Bilan des mesures sur les données anglaises 127

2. Données naxi 128
2.1. Matériau linguistique 128
2.2. Réflexion sur les phrases-cadres 132
2.2.1. Phrases-cadres et contexte d'élicitation pour les quatre premiers locuteurs (F2, M5, M7, M9) 132
2.2.2. Conditions spéciales pour le locuteur M8 133
2.3. Bilan des enregistrements 133
2.4. Représentation visuelle des données 134
2.4.1. F_0 et quotient ouvert 134
2.4.2. Formants 134
2.4.2.1. Comparaison entre tons lexicaux 136
2.4.2.2. Comparaison entre conditions de lecture 136
2.5. Tableaux de résultats 137
2.5.1. Les valeurs moyennes de F_0 augmentent en lecture insistante 137
2.5.2. Evolution de F_0 : variété des stratégies 137
2.5.3. Quotient ouvert : cohérence des stratégies individuelles 138
2.5.4. La longueur de la rime varie peu 139
2.5.5. L'intensité acoustique globale est plus élevée en lecture insistante, mais la différence est moindre que dans les données anglaises 140
2.5.6. Bilan des différences entre conditions de lecture dans les données naxi 141

3. Données vietnamiennes 143
Rappel de l'enjeu scientifique de la prise en compte du vietnamien : 143
3.1. Matériau linguistique 144
3.2. Réflexion sur les phrases-cadres 146
3.3. Bilan des enregistrements et résultats 148
3.3.1. Analyse qualitative 149
3.3.1.1. Ton B2 149
Reprise du voisement après la constriction glottale dans environ 2 cas sur 5 en condition de lecture insistante 151
3.3.1.2. Ton D2 151
3.3.1.3. Ton D1 153
3.3.2. Analyse quantitative 154
3.3.2.1. Longueur des cycles glottiques (fréquence fondamentale) : description du contraste entre tons et entre conditions de lecture 154
3.3.2.2. Quotient ouvert : les valeurs des tons B2 et D2 contrastent de façon maximale ; elles varient peu en fonction de la condition de lecture 157
3.3.2.3. Stratégies opposées dans l'emploi intonatif du paramètre de longueur 158
3.3.2.4. Intensité acoustique globale (mesure réalisée sur une partie du corpus seulement) 159
3.3.2.5. L'hypothèse d'une corrélation entre DECPA et débit d'air 159
3.3.3. Bilan : les caractéristiques de qualité de voix ressortent plus nettement encore en lecture insistante (constriction glottale du ton B2, phonation modale des tons D1 et D2) 160

4. Synthèse des résultats : comparaison entre les trois langues 161

4.1. Importance des modifications locales 161
4.2. Relation entre fréquence fondamentale et intensité dans les trois langues étudiées 162
4.2.1. En anglais : fortes modifications de F_0 et de l'intensité ; en naxi et vietnamien : modification relativement plus élevée de l'intensité 162
4.2.2. Le rapport $\Delta I / \Delta F_0$ est plus bas lors d'un changement de ton lexical que lors d'un changement de lecture soignée à lecture insistante 163

Chapitre IV. Discussion générale 167

1. L'hypothèse d'un partage des ressources entre accentuation et intonation 167
Mise au point de vocabulaire : intensification intonative et accent d'insistance, marquage intonatif du focus, et focalisation contrastive 169
1.1. La qualité de voix (mesurée par le quotient ouvert) varie moins en vietnamien, langue dans laquelle elle participe à une opposition lexicale 170
1.2. Les rimes connaissent un allongement considérable en naxi, langue dans laquelle le trait de longueur n'est pas distinctif 170
1.3. Les langues à tons contrôlent l'emploi intonatif de F_0 : le ratio entre variation de F_0 et variation d'intensité est plus bas que dans les données anglaises 171
1.3.1. Le problème de la mise en valeur prosodique d'une syllabe portant un ton bas ou moyen, et une solution catégorielle 171
1.3.2. La variation de F_0 lors d'un surcroît d'intensification intonative diffère d'une langue à l'autre 173
1.3.3. Eléments de réflexion sur les mécanismes physiologiques sous-jacents à l'effort vocal 173
1.3.4. Les résultats d'études sur d'autres langues paraissent confirmer l'hypothèse d'un partage des ressources entre accentuation lexicale et intonation 175
1.4. La variation des fréquences formantiques entre lecture soignée et lecture insistante est non significative dans la très grande majorité des cas 176

2. Implications théoriques des observations : choix pour une modélisation prosodique 178
2.1. Présentation de l'approche autosegmentale-métrique, et comparaison de l'emploi de la notion de ton dans les divers modèles 178
2.1.1. Approche autosegmentale-métrique et traditions linguistiques américaines 179
2.1.2. Influence des modèles de tonologie africaine sur l'approche autosegmentale-métrique 180
2.1.3. Lien entre le développement de l'approche autosegmentale-métrique et les besoins constatés dans le domaine du traitement automatique des langues 180
2.1.4. La notion de ton dans les modèles autosegmentaux-métriques et dans d'autres modèles (« britannique », « superpositionnel », « morphologique ») : un même mot, des emplois qui divergent 180
2.1.5. Une version radicale du postulat des tons intonatifs : l'hypothèse d'une identité structurelle entre prosodie de langues à tons lexicaux et prosodie de langues sans tons lexicaux 181
2.2. Le difficile débat entre approche autosegmentale-métrique, approche superpositionnelle, approche britannique et approche morphologique 183
2.2.1. Principaux éléments du débat 183
2.2.2. Résultats expérimentaux qui ne paraissent pas réductibles à une approche en termes de séquences tonales 186
2.2.2.1. Observations sur le marquage des frontières en naxi 186
2.2.2.2. Divergence entre nos observations et celles de Pierrehumbert et Liberman sur l'insistance en anglais 187
2.2.2.3. Bilan 188
2.2.3. Réflexions sur l'idée d'une superposition 188
2.2.4. Possibilités d'application didactique des modèles superpositionnels 190

3. Réflexion typologique 191
3.1. Quelques prototypes de systèmes prosodiques 192
3.1.1. Accent libre et accent fixe 192
3.1.2. Absence de phénomènes d'accentuation lexicale (« langues sans tons et sans accent de mot ») 194
3.1.3. Systèmes de tons ponctuels 194
3.1.3.1. Systèmes prosodiques africains à tons ponctuels 195
3.1.3.2. Systèmes prosodiques asiatiques à tons ponctuels : implications typologiques de la description des langues à tons ponctuels d'Asie 196
3.1.4. Registres de qualité de voix 198
3.1.5. Tons modulés 201
3.1.5.1. Le débat sur l'existence de tons modulés. Les propositions de réanalyse des tons modulés en tons ponctuels 201
3.1.5.2. Apparition et évolution des tons modulés 204
3.1.5.3. Un exemple développé de langue à tons modulés et qualité de voix : le vietnamien 204
3.1.5.4. Le domaine du ton n'est pas nécessairement la syllabe 206
3.1.6. Entre système tonal et système accentuel : la question de l'accent dans les systèmes tonals, et les systèmes à accent mélodique (pitch accent) 208
3.1.6.1. Problématique générale : tons et accent 208
3.1.6.2. Les langues à accent mélodique forment-elles un pôle typologique distinct ? 211
3.1.6.2. Une libre combinaison des tons et d'un accent lexical est-elle possible ? L'exemple du ma'ya 213
3.2. Une nouvelle perspective typologique, indépendante de la présence ou de l'absence de tons lexicaux : langues à *prosodie calculée* et langues à *prosodie contrôlée par l'énonciateur* 214
3.2.1. Intuition de départ 216
3.2.2. Formulation proposée par Larry Hyman : tons paradigmatiques et tons syntagmatiques 218
3.2.3. Le sandhi est plutôt une caractéristique d'une langue à prosodie contrôlée, le downstep plutôt une caractéristique d'une langue à prosodie calculée 220
3.2.4. Choix terminologique 221

Conclusion *225*

1. Conclusions au plan de la technique expérimentale 225
1.1. Possibilité d'une évaluation du quotient ouvert par électroglottographie, même dans les cas de phonation non modale 225
1.2. Le quotient ouvert n'est pas déductible de F_0 225
1.3. Utilité de la mesure d'intensité acoustique globale dans la comparaison des mêmes mots dans différents contextes d'énonciation 226

2. Conclusions concernant la prosodie des trois langues étudiées 226
2.1. Conclusions au sujet du vietnamien : données nouvelles sur la qualité de voix associée aux tons, et perspectives typologiques 226
2.2. Conclusions au sujet du naxi 227
Le naxi n'utilise pas la qualité de voix de façon lexicalement distinctive 227
Tons flottants, réduction syllabique et contours intonatifs dans la prosodie naxi 227
2.3. Conclusions au sujet de l'anglais 228
2.4. Utilité de la comparaison entre langues 228

3. Conclusions au plan de la modélisation prosodique et de la typologie 229
3.1. Confirmation de l'hypothèse d'un partage des ressources entre accentuation et intonation 229

3.2. Variété typologique. Proposition de deux pôles typologiques 229
3.2.1. Réflexions au sujet des modèles autosegmentaux-métriques de l'intonation 230
3.2.2. Une perspective typologique nouvelle : langues à prosodie calculée et langues à prosodie contrôlée par l'énonciateur : différences typologiques dans le degré de contrôle de la prosodie par l'énonciateur 230

Glossaire et explication des abréviations *233*

Accent 233
Accent contrastif ; accent de focalisation ; accent de phrase 234
Accent d'insistance 234
Accentuation 235
AS 235
Consultant 236
Craquée (voix craquée) 236
Déclinaison 236
DECPA 236
Downstep 237
Electroglottographie 237
F_0 237
F_1, F_2, F_3 237
FK 237
Focalisation contrastive 237
Focus 238
Glottalisation 238
Groupement intonatif 239
H, M, L : abréviation de ton haut, ton moyen, ton bas 239
Insistant (condition I : condition de lecture insistante) 239
Intensification intonative 240
Intonation 240
Intonème 240
L : abréviation de Low (ton bas) 240
Laryngalisation 241
M : abréviation de Moyen (ton moyen). 241
Marquage intonatif du focus 241
Mécanisme phonatoire 241
Mélodie 242
Mise en valeur intonative 243
Morphème intonatif 243
NL 244
O_q : abréviation de quotient ouvert (voir ce terme) 244
Paramètre 244
Prosodie 244
Qualité de voix 245
Quotient ouvert (indication sur la qualité de voix) 245
Registre (langues à registres de qualité de voix) 246
Segmental et suprasegmental 246
Soigné (condition S = condition de lecture soignée) 246
Ton, ton ponctuel, ton modulé 246
Tonème 248

Références *249*

Annexe 1. Observations sur les phonèmes du naxi 297

0.1. Brève présentation du naxi. La question de sa position vis-à-vis du groupe birman-yi (« birman-ni », « lolo-birman ») 297

0.2. Variété dialectale. Parlers étudiés 298

0.3. Principaux enjeux de la description phonémique du naxi 300

1. Les initiales 305

1.1. L'initiale palatale /ç/, fruit d'anciennes initiales nasales ? Oppositions de nasalité sur les voyelles et oppositions de consonnes fricatives 306

1.1.1. Description et mise en correspondance entre dialectes 306

1.1.2. Premières hypothèses diachroniques sur l'origine des voyelles nasales 310

1.1.3. Données complémentaires pour la suite de la réflexion 311

1.2. Allophones rétroflexes des consonnes coronales, et rétroflexion phonémique 312

1.2.1. Description 312

1.2.2. Analyse synchronique 315

1.2.3. Comparaison entre dialectes 316

1.3. Allophones palatalisés des vélaires devant /i/, /y/, /jɤ/, /ja/, /jɑ/, /jə/ 318

1.3.1. Description générale du phénomène 318

1.3.2. Rôle de la langue nationale dans la palatalisation 319

1.3.3. La palatalisation a-t-elle abouti à la perte d'oppositions syllabiques ? 320

1.3.4. Choix de conserver un statut distinct à certains /ɯ/ modifiés 323

1.4. Autres observations concernant les initiales 324

1.4.1. Allophones « trills » des occlusives bilabiales /p^h/, /p/, /b/, /m͡b/ devant /ɣ/ 324

1.4.2. Allophones uvulaires des vélaires devant les voyelles d'arrière 325

1.4.3. Fricatives 325

1.4.4. Nasales 326

1.4.5. Latérale l, et confusions avec n͡d 327

1.4.6. Occlusives et affriquées : description en quatre degrés de *Voice Onset Time*, et recoupements possibles avec l'opposition *fortis/lenis* 327

Comparaison entre les occlusives naxi et françaises : occlusives non voisées et occlusives fortis 328

Perte fréquente du voisement pendant l'occlusion des occlusives et affriquées prénasalisées 329

2. Observations synchroniques sur les rimes 329

2.1. Syllabes labialisées : rime /wə/, initiales labialisées, ou médiale /w/ ? 330

2.2. La fricative /ɣ/, noyau de syllabe 333

Proximités phonétiques de /ɣ/ 334

2.3. Notation des voyelles d'aperture moyenne 335

2.4. Réalisation phonétique des voyelles d'arrière non arrondies 336

2.5. Notation des voyelles ouvertes 337

2.6. La question de la « voyelle neutre » 337

2.7. Attaque des voyelles syllabiques 338

2.8. Possibilité de fusion entre une syllabe sans consonne initiale et la syllabe qui précède 340

2.9. Semi-voyelles 342

2.10. La voyelle rhotique : confirmation de l'absence d'oppositions de qualité de voix en naxi 344

3. Prononciation des phonèmes de la variété locale de chinois 345

3.1. Remarques générales 345

3.1.1. Situation actuelle de la langue par rapport au chinois 345
3.2. La langue chinoise parlée par l'informateur M4 348
3.2.1. Maintien intermittent d'une distance phonétique entre naxi et chinois 348
3.2.2. Suppression des nasales finales, et phénomènes d'hyper-correction 350
3.2.3. Conservation des oppositions portées par les nasales lorsqu'elles sont en position initiale 350
3.2.4. Changements dont le « chinois deuxième langue » est potentiellement porteur 351
3.2.4.1. Changements dans les initiales : dans les mots de « chinois deuxième langue », /v/ a statut d'initiale, la combinaison /ni/ est très fréquente, /ŋ/ peut former syllabe 351
3.2.4.2. Changements dans les rimes : développement de rimes à semi-voyelle 351
3.2.4.3. Changements dans les tons 353
Brève note sur les changements syntaxiques 357
4. Phénomènes à l'échelle du mot : fusion syllabique et intégration nominale 357
4.1. Fusion syllabique 357
4.2. Phénomènes d'intégration nominale 358
4.2.1. Modifications tonales 358
4.2.1.1. Exposition des faits 358
4.2.1.2. Hypothèses sur l'origine diachronique du changement tonal 361
4.2.2. Modifications de phonèmes 363
5. Variations intonatives dans la réalisation des phonèmes 365
6. Aperçu de la variation dialectale : FK, NL 367
6.1. Dialecte de FK (locuteur M12) 367
6.1.1. Les initiales 367
6.1.1.1. Observations concernant les initiales coronales 367
6.1.1.2. Observations concernant les consonnes bilabiales 368
6.1.2. Les rimes 368
6.1.2.1. Les rimes simples 368
6.1.2.2. Les semi-voyelles 368
6.1.3. Les tons 369
6.1.3.1. Quatre tons lexicaux, plus un ton sur les emprunts chinois 369
6.1.3.2. Tons flottants 369
6.1.3.3. Variation expressive des tons 370
6.1.3.4. Phénomènes de sandhi 371
6.2. Dialecte de NL (locuteur M10) 372
6.2.1. Les rimes 372
6.2.2. Les initiales 373
6.2.3. Les tons 373
6.3. Eléments de comparaison entre dialectes 373
6.3.1. Généralités 374
6.3.2. Distinctions perdues et confusions en cours dans le dialecte de DYZ et autour de la ville de Lijiang 376
6.3.2.1. Affriquées et fricatives rétroflexes et non rétroflexes 376
6.3.2.2. Occlusives et affriquées prénasalisées et voisées 377
6.4. Vérifications à effectuer dans le dialecte de Zhong^{1}He^{2} 中和, décrit par Fu Maoji 1981 377
Eléments de conclusion 378

Annexe 2. Résultats expérimentaux complémentaires, et outils informatiques 381

1. Données anglaises complémentaires 381

1.1. Résultats des mesures sur les données de l'expérience-pilote avec le locuteur anglais M1 381

1.2. Réalisation des consonnes initiales *lenis* et *fortis* 383

1.3. Nécessité d'exclure les consonnes continues voisées initiales de syllabe 383

1.3.1. Problématique générale : la division de la syllabe en initiale et rime 384

1.3.2. Arguments fournis par les données anglaises 385

1.4. Caractéristiques intrinsèques des phonèmes vocaliques 389

2. Résultats de mesures de débit d'air des tons vietnamiens B2, D1 et D2 391

2.1. Déroulement de l'enregistrement 392

2.2. Pose des bornes de début et fin 396

2.3. Résultats de débit d'air oral 396

3. Etude de DECPA, amplitude des pics de fermeture sur le signal électroglottographique 399

3.1. Introduction : la question de la mesure d'intensité, et l'utilité de mesures de la pente spectrale 400

3.2. A quoi correspond le paramètre DECPA ? 402

4. Scripts développés dans l'environnement MATLAB 404

4.1. Outil d'analyse du signal électroglottographique : implémentation logicielle de la méthode de détection des pics positifs et négatifs sur la dérivée du signal 405

4.1.1. Programmes existants au début de notre recherche 405

4.1.1.1. Programmes de Nathalie Henrich 405

4.1.1.2. Programmes de Vu-Ngoc Tuân 406

4.1.2. Orientations du développement logiciel effectué au cours de la présente étude 408

4.1.2.1. Choix d'une méthode semi-automatique 408

4.1.2.2. Choix de ne pas appliquer de traitement sur le signal enregistré 408

4.1.3. Principes pour la détection des pics de fermeture 409

Traitement des pics de fermeture doubles ou multiples 410

4.1.4. Principes pour la détection des pics d'ouverture 410

1.4.1. Choix de lisser le signal 411

1.4.2. Détection du minimum local ou barycentre des pics : Choix visuel au vu de quatre résultats 412

4.1.5. Réflexion sur les méthodes de détection des pics de fermeture pour le calcul de la fréquence fondamentale. Proposition de deux algorithmes. 415

Conclusion au sujet de la fréquence fondamentale 417

4.2. Interface de vérification de valeurs de formants 418

4.3. Implémentation d'un test-t apparié 418

4.4. Simulations syllabiques 419

4.4.1. Explication au sujet du nombre de points des courbes rééchantillonnées 419

4.4.2. Traitement des cas où certains points de données manquent 419

4.4.3. Des moyennes aux simulations syllabiques 420

4.4.4. Perspectives de développement futur 421

Annexe 3. Perspectives théoriques complémentaires 423

1. L'alignement des points-clefs de la courbe de fréquence fondamentale avec la « ligne des segments » : la problématique et sa remise en cause 423

1.1. Le déplacement du pic de F_0 correspondant à l'« accent de phrase » a un effet sur l'interprétation de l'énoncé 424

1.2. Limites de la problématique : l'alignement est-il un paramètre indépendant des autres paramètres phonétiques ? 426

1.2.1. La fréquence fondamentale n'est pas un paramètre indépendant des autres paramètres phonétiques 427

1.2.1.1. Existence de relations d'équivalence et de compensation entre paramètres 427

1.2.1.2. Des données expérimentales établissent le rôle de l'intensité et de la pente de F_0 dans la perception des contrastes d'alignement du pic de F_0 427

1.2.2. Le mouvement des articulateurs, comme la courbe de F_0, est continu 428

2. Tons et structure métrique en chinois 430

Figures du volume I 435

Graphiques 491

Figures des Annexes 494

Cartes 514

Index 516

Table des figures du volume I 518

Table des photographies et des cartes 520

Table des graphiques 520

Table des figures des Annexes 521