



**HAL**  
open science

# Analyse de systèmes prosodiques d'Asie orientale : linguistique de terrain et phonétique expérimentale

Alexis Michaud

► **To cite this version:**

Alexis Michaud. Analyse de systèmes prosodiques d'Asie orientale : linguistique de terrain et phonétique expérimentale. Linguistique. Université Lumière - Lyon 2, 2017. tel-01538660v1

**HAL Id: tel-01538660**

**<https://theses.hal.science/tel-01538660v1>**

Submitted on 13 Jun 2017 (v1), last revised 31 Jul 2017 (v3)

**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.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

# Université Lumière - Lyon 2

Mémoire de synthèse en vue de l'Habilitation à diriger des recherches

---

## Analyse de systèmes prosodiques d'Asie orientale : linguistique de terrain et phonétique expérimentale

---

Alexis MICHAUD

Sous la direction de M. François PELLEGRINO

Présentée et soutenue publiquement le 12 juin 2017

### Composition du jury :

- Mme le Professeur Juliette BLEVINS, City University of New York
- M. le Professeur Denis CREISSELS, Université Lumière – Lyon 2
- M. Guillaume JACQUES, CNRS / INALCO
- Mme le Professeur Wai-Sum Vanti LEE, City University of Hong Kong
- M. le Professeur Stephen MOREY, La Trobe University
- M. François PELLEGRINO, CNRS / Université Lumière – Lyon 2 (*coordinateur*)
- Mme le Professeur Jacqueline VAISSIÈRE, Université Sorbonne Nouvelle – Paris 3

Die Anregung zur Bildung des Ichideals, als dessen Wächter das Gewissen bestellt ist, war nämlich von dem durch die Stimme vermittelten kritischen Einfluß der Eltern ausgegangen, an welche sich im Laufe der Zeiten die Erzieher, Lehrer und als unübersehbarer, unbestimmbarer Schwarm alle anderen Personen des Milieus angeschlossen hatten. (Die Mitmenschen, die öffentliche Meinung)

L'incitation à la formation du moi idéal, dont la conscience a été préposée gardienne, est née de l'influence critique des parents, médiée par la voix. Parents à qui se sont joints au fil du temps éducateurs et enseignants, et l'essaim sans fin des autres personnes qui composent un société. (*Les autres*, l'opinion publique.)

S. Freud, *Zur Einführung des Narzißmus* (1924)

Ce mémoire en vue de l'Habilitation à diriger les recherches a été préparé à l'Université Lumière - Lyon 2. Les centres de recherche qui m'ont accueilli dans le cadre de mes travaux de 2006 à 2016 sont :

- Le laboratoire de **Langues et civilisations à tradition orale** (en abrégé Lacito), unité mixte de recherche (UMR 7107) sous tutelle du Centre National de la Recherche Scientifique, de l'Université Sorbonne Nouvelle – Paris 3 et de l'Institut national des langues et civilisations orientales. (Périodes : octobre 2006 à décembre 2010, et de juillet 2016 jusqu'à maintenant.)
- Le **Centre d'études français sur la Chine contemporaine** (en abrégé CEFC), unité mixte des Instituts Français de Recherche à l'Etranger (UMIFRE 18) sous tutelle du Ministère des Affaires Etrangères, du Centre National de la Recherche Scientifique et de l'Institut Français de Recherche à l'Etranger. (Période : janvier 2011 à octobre 2012.)
- L'Institut international de recherche **Multimédia, Information, Communication et Applications** (en abrégé MICA), institut de recherche franco-vietnamien (UMI 2954) sous tutelle de l'Institut Polytechnique de Hanoi (HUST), du Centre National de la Recherche Scientifique et de Grenoble INP. (Période : novembre 2012 à juin 2016.)

Je tiens à remercier les membres du jury de leur participation. La liste des collègues qui m'ont aidé au fil des ans est si longue qu'on me permettra de les remercier collectivement, et de ne nommer – tout subjectivement – que quelques complices au sein des centres de recherche qui m'ont accueilli. Au laboratoire de Langues et Civilisations à Tradition Orale, mes remerciements particuliers à Evangelia Adamou, Alex François, Zlatka Guentchéva, Séverine Guillaume, Martine Mazaudon, Boyd Michailovsky, Lameen Souag. Au Vietnam, à Eric Castelli, Hằng Đình, Ngọc-Diệp Đỗ, Đăng-Khoa Mạc, Việt-Sơn Nguyễn, Hiễn Phạm, Đỗ-Đạt Trần. Au Centre d'études français sur la Chine contemporaine, à Jean-François Huchet, Paul Jobin et Stéphane Corcuff. Au Laboratoire de Phonétique et Phonologie, à Angélique Amelot, Didier Demolin, Cécile Fougeron, Cédric Gendrot, Barbara Kühnert, Rachid Ridouane.

Enfin, je remercie, pour nos apprentissages partagés, les étudiants-chercheurs dont j'accompagne les travaux à divers titres.

# Sommaire

Introduction.....	5
1. Travaux réalisés .....	10
1.1. Linguistique de terrain.....	10
1.1.1. Etude de la langue na de Yongning.....	11
1.1.2. Etude d'autres parlers du groupe naish .....	16
1.2. Phonétique expérimentale et phonologie théorique .....	21
1.2.1. Phonétique expérimentale.....	21
1.2.2. Phonologie théorique.....	25
1.3. Prosodie : études de cas et réflexions générales.....	27
1.3.1. Etudes de cas.....	27
1.3.2. Réflexions générales et typologie prosodique.....	28
1.4. Phonétique historique.....	31
1.4.1. Etude de la phonologie historique des langues naish.....	32
1.4.2. Etude du transfert de la nasalité entre un groupe de consonnes initial et une voyelle	33
1.4.3. Autres travaux en phonétique historique.....	34
1.5. Méthodologie : de la collecte à l'archivage des données linguistiques.....	36
1.5.1. Méthodes de collecte et analyse de données .....	37
1.5.2. Archivage et diffusion des données.....	43
2. Orientations futures .....	53
2.1. Perspective aréale .....	54
2.1.1. Domaine des langues naish (sino-tibétain).....	54
2.1.2. Domaine des langues vietiques (austro-asiatique).....	57
2.2. Perspective thématique : la « linguistique des langues » à l'ère numérique .....	58
2.2.1. Tâches fondamentales pour la création de données connectées .....	59
2.2.2. Vers une modélisation de haut niveau .....	61
Conclusion .....	62
Références citées .....	65
Table des matières.....	74

# Introduction

Dans le prolongement d'une thèse intitulée *Prosodie de langues à tons (naxi et vietnamien), prosodie de l'anglais : éclairages croisés* (2005), j'ai proposé lors de ma candidature au Centre National de la Recherche Scientifique en 2006 un programme de recherche intitulé « Contribution à la typologie prosodique, sur la base d'enquêtes de première main ». Dans l'espoir d'emporter l'adhésion du Comité national chargé des recrutements, le projet promettait d'allier plusieurs approches : non seulement phonétique expérimentale et modélisation phonologique, dans la tradition du laboratoire où j'avais préparé ma thèse (le Laboratoire de Phonétique et Phonologie de Paris), mais aussi phonétique historique et typologie. Une décennie après, la candidature à l'Habilitation à diriger les recherches offre l'occasion d'un bilan d'étape.

Conformément au titre du programme de 2006, prosodie et typologie tiennent une grande place dans mes travaux, de même que les enquêtes linguistiques de première main. Mais le contexte du présent mémoire de synthèse est différent de celui des concours de recrutement : il autorise une expression allégée des considérations tactiques que ne peut ignorer le candidat qui se présente devant un jury de recrutement. Le projet de 2006 présentait les enquêtes de terrain comme le moyen de parvenir à apporter une « *Contribution à la typologie prosodique* » : autrement dit, la typologie était présentée comme la finalité des études de cas. Cette formulation était choisie dans l'idée de rassurer ceux parmi les examinateurs qu'aurait inquiétés l'annonce d'une plongée sans retour dans la diversité des terrains, des langues et des données : les enquêtes de première main (fondement de la linguistique dite « de terrain ») étaient d'emblée inscrites dans un trajet *du terrain à la théorie*, par lequel elles avaient vocation à nourrir un modèle typologique, à vocation universelle<sup>1</sup>.

---

<sup>1</sup> La formule « Du terrain à la théorie » est le titre du colloque anniversaire des quarante ans du laboratoire LACITO, en 2016. Un ouvrage collectif du même laboratoire s'intitulait « Du terrain au cognitif », suivant un argumentaire semblable, dans lequel on fait valoir les enquêtes de terrain comme point de départ d'un trajet dont l'élaboration théorique la plus poussée constitue le point d'aboutissement.

Comme l'écrit Alexandre François en avant-propos d'une étude de la langue mwotlap (Vanuatu) :

J'avais pour dessein d'aller trouver aux antipodes des configurations inédites, à la fois uniques au monde, et pourtant susceptibles de me faire entrevoir les contours de l'universel. (François 2003:xvii)

La mise en perspective typologique constitue assurément un outil précieux au cours de l'analyse d'une langue particulière ; et chaque monographie consacrée à une langue a vocation à contribuer à l'enrichissement des modèles typologiques. C'est là un argument que les explorateurs de la diversité des langues savent rappeler à leurs collègues phonéticiens pour justifier l'intérêt porté aux « petites langues » :

(...) languages that are in danger of disappearing are of extreme importance; without a good sense of how languages vary, not only in terms of the symbolic units such as phonemes and allophones but also in the details of their phonetic implementation, we have little hope of understanding the possible range of language. Endangered languages in particular represent important but often-ignored source of information about what is possible in Language. (Richard Wright, directeur du laboratoire de phonétique de l'Université de Washington)<sup>2</sup>

Pour autant, ce qui anime mon travail est d'abord un intérêt pour la réalité de telle langue dans son fonctionnement : considérer d'abord *la langue comme système*, pourrait-on dire, pour autant que cette formule (qui prête bien sûr à diverses interprétations et divers malentendus) aide à cerner un dénominateur commun à l'enseignement d'auteurs comme Saussure, Trubetzkoy, Martinet, Benveniste, et (pour ne citer que des auteurs francophones) parmi ceux que j'ai eu le plaisir d'entendre de mes oreilles, Michel Launey, Antoine Culioli, Denis Creissels... S'intéresser aux faits de langues en système, cela ne signifie pas que la langue soit idéalisée, comme un tout dont chaque élément serait indissolublement lié à tous les autres. Cela signifie que les diverses propriétés d'une langue entretiennent des liens dissimulés dont doit nécessairement faire abstraction la typologie, même la plus pénétrante (sans parler d'approches typologiques « à coups de serpe », qui prélèvent trop rapidement, dans une langue mal connue, de petits ensembles de données pour sustenter une activité théorisante). En ce sens, la linguistique dite « de terrain » n'est pas intégralement soluble dans la typologie. Les recueils de textes, les dictionnaires et les grammaires de langues peu documentées constituent des trésors irremplaçables, dont la valeur ne se limite pas aux arguments qu'ils peuvent livrer en faveur de telle ou telle proposition typologique, de tel ou tel modèle linguistique.

---

<sup>2</sup> Source : site internet de l'Université de Washington, <http://depts.washington.edu/phonlab/people/wright.htm> consulté en 2015.

Dans cette perspective, le titre adopté ici, « Analyse de systèmes prosodiques d'Asie orientale : linguistique de terrain et phonétique expérimentale », revient à présenter l'analyse du système prosodique d'une langue comme une tâche à part entière, et presque une fin en soi. La visée typologique du travail ne disparaît pas, mais sa présence est désormais en filigrane, dans le pluriel des *systèmes prosodiques* étudiés : si ces systèmes peuvent être réunis sous une appellation commune, c'est parce qu'ils ne sont pas incommensurables les uns aux autres, et peuvent s'éclairer réciproquement. Autrement dit, dans la recherche constante d'un équilibre entre les deux aspects – tous deux indispensables – que constituent la réflexion en linguistique générale et l'examen de faits particuliers, j'ai eu à cœur, pendant la période concernée (2006-2016), d'accorder une attention particulière aux fondations empiriques de la recherche<sup>3</sup>. Les recherches au sujet de parlars jusque-là peu connus demandent un effort d'apprentissage de la langue et d'établissement des faits ; je me suis efforcé d'accorder à cet aspect du travail l'attention qu'il demandait. Ce choix est en phase avec l'environnement scientifique du laboratoire Lacito (mon rattachement institutionnel « pérenne » depuis 2006), un laboratoire *à l'écoute des langues*, où se pratique un égalitarisme militant entre langues, en vertu duquel la plus « petite » des langues en termes de nombre de locuteurs, de prestige, d'extension géographique ou de documentation existante est considérée comme aussi digne d'intérêt que les langues riches d'une longue tradition écrite.

Outre la typologie, j'ai également choisi de maintenir à l'arrière-plan dans le présent mémoire un second type de modélisation fondé sur la comparaison entre langues : l'approche *panchronique* au sens d'Haudricourt (Haudricourt 1940; Haudricourt 1973; Haudricourt 1978), proche de ce que Juliette Blevins nomme « Evolutionary Phonology » (Blevins 2004; Blevins 2006), expression qu'on traduira ici par « phonologie évolutionniste ». Il s'agit d'un programme de recherche ambitieux : une typologie du changement linguistique qui révèle les raisons d'apparition et de disparition de phénomènes linguistiques. L'approche panchronique en phonologie vise à parvenir à des lois du changement phonétique qui atteignent le plus haut degré de généralité : non pas formuler de simples constats au sujet de changements intervenus dans telle langue (sens dans lequel le mot « loi » est communément employé en linguistique historique : « loi de Grimm », « loi de Werner »... constituent des constats au sujet d'une langue – ou d'un groupe de langues –

---

<sup>3</sup> En cela, mes travaux sont en phase avec le projet scientifique du « Laboratoire d'excellence » (2011-2019) « Fondements empiriques de la linguistique / Empirical Foundations of Linguistics » (EFL).



en particulier), mais dégager le conditionnement précis d'un changement, et par là parvenir à des lois indépendantes de l'espace et du temps. De même, la phonologie évolutionniste vise à déceler dans l'organisation des faits synchroniques la trace des processus par lesquels ils sont apparus.

Evolutionary Phonology is the study of synchronic sound patterns as partial reflections of their evolution or history. Central to Evolutionary Phonology is the attempt to explain relationships between sound patterns and sound change, and, more generally, to explain why sound patterns have the typological distributions they do. (Juliette Blevins, exposé « Sound patterns and sound change: New threads in the panchronic tapestry » au colloque « Du terrain à la théorie : Les 40 ans du Lacito », 16 novembre 2016)

Les approches dynamiques que constituent la phonologie panchronique et la phonologie évolutionniste posent un niveau d'exigence élevé dans la modélisation des phénomènes, et encouragent à pousser toujours plus avant la recherche des causes du changement. Mais la première synthèse tentée par Hagège en collaboration avec Haudricourt (Hagège & Haudricourt 1978) a quelque chose d'inabouti : l'ouvrage reste à un stade programmatique, sans réellement poser les fondements de cette nouvelle approche de la phonologie. La difficulté de l'entreprise panchronique tient à la multiplicité des facteurs susceptibles d'exercer une influence sur l'évolution d'une langue. A l'époque où le projet a été formulé, la modélisation d'un vaste ensemble de facteurs pouvait paraître hors d'atteinte : ainsi, Martinet était tenté de se détourner de l'entreprise, à laquelle il reconnaissait pourtant une valeur heuristique.

Contrary to some other functionalists, I am not tempted to posit panchronic laws of phonological evolution. But looking for general laws may lead to what I would call the positing of useful hypotheses. (Martinet 1996:164)

Une analogie avec la météorologie peut être éclairante : les phénomènes atmosphériques tels que les nuages, les précipitations ou le vent sont régis par des lois dont chacune prise en particulier est simple. L'air chaud a tendance à s'élever ; ce faisant, il se dilate et sa température diminue, ce qui occasionne la condensation de la vapeur d'eau, à saturation, en gouttelettes. Celles-ci forment un nuage ; elles fusionnent pour donner des précipitations liquides ou solides selon des paramètres bien identifiés. Pour autant, le nombre de paramètres en jeu et la combinatoire de leurs effets sont tels que la modélisation météorologique doit faire appel à diverses disciplines : outre la mécanique des fluides et la thermodynamique, elle recourt aux outils de diverses branches de la physique, de la chimie et des mathématiques. S'il est impossible de connaître avec certitude le temps qu'il fera dans quelques semaines, les prévisions météo contemporaines permettent de connaître le

temps qu'il fera dans les jours qui viennent. De même, la langue est un objet d'une complexité multi-dimensionnelle (Pellegrino et al. 2009), et le changement linguistique appelle une modélisation aussi soignée que les phénomènes météorologiques.

Le projet d'une approche dynamique des systèmes phonologiques (et, plus généralement, des phénomènes linguistiques) peut aujourd'hui bénéficier d'outils informatiques qui permettent la prise en compte de facteurs multiples. J'aimerais apporter une contribution à ce domaine à moyen ou long terme, mais force est de constater que mes efforts dans ce domaine restent jusqu'à maintenant bien modestes : sans commune mesure avec l'ampleur du programme scientifique. C'est pourquoi, tout en gardant à l'esprit le projet d'une approche évolutionniste et panchronique des phénomènes, je me suis abstenu de faire figurer cette entreprise dans le titre de la présente synthèse.

Ce mémoire est classiquement divisé en un volet concernant les travaux réalisés et un volet prospectif. Dans le premier volet sont abordées les questions de méthode de collecte, analyse et archivage de données linguistiques (§1.5), questions qui me paraissent mériter d'être discutées en détail, en veillant bien sûr à ne pas tomber dans « le 'méthodologisme' qui consiste à essuyer sans cesse ses lunettes au point d'oublier de voir dedans » (formule empruntée au psychanalyste Paul-Laurent Assoun)<sup>4</sup>. Parmi les travaux réalisés, neuf publications ont été versées au dossier ; un encadré en tête de sous-chapitre fournit la liste de celles qui sont les plus pertinentes pour le thème en question.

---

<sup>4</sup> *Psychanalyse*, Paris : Presses Universitaires de France, 2007, p. XVIII.

# 1. Travaux réalisés

## 1.1. Linguistique de terrain

Documents joints au dossier concernant ce volet :

- 1 ouvrage sous presse :

*Tone in Yongning Na: lexical tones and morphotonology*. (Studies in Diversity Linguistics). Berlin: Language Science Press (2017).

- 1 article :

Tonal reassociation and rising tonal contours in Naxi. *Linguistics of the Tibeto-Burman Area* 29(1). 61–94 (2006).

Entre 2006 et 2016, les enquêtes linguistiques ont tenu une place centrale dans mon travail. Depuis que j'avais appris l'existence de la linguistique de terrain (en 1994), j'avais le souhait de me consacrer à l'étude d'une langue jusque-là inconnue ou peu décrite. Au cours de mes années de thèse (2002-2005), j'avais connu une certaine expérience des enquêtes linguistiques sur le terrain : en 2002 puis 2004, je m'étais rendu dans le Yunnan (Chine) pour y recueillir des données de langue naxi (une langue sino-tibétaine). Mais j'étais alors absorbé par le projet de mener à terme dans le temps imparti (trois années) une thèse qui combinait phonétique expérimentale et typologie linguistique, tout en satisfaisant aux diverses obligations d'enseignement et de formation à la recherche des « doctorants-moniteurs ». Cette charge de travail ne s'est pas avérée compatible avec les tâches de patience par lesquelles s'acquiert la connaissance approfondie d'une langue, et s'élabore une description linguistique digne de ce nom. Mettre au propre les récits transcrits sur le terrain, toiletter et enrichir une liste de mots pour qu'elle devienne un lexique puis un dictionnaire, et décrire les mécanismes rencontrés dans la langue-cible : il s'agit là d'un travail qui nécessite une bonne maîtrise de la langue concernée.

Mon entrée au CNRS a constitué un tournant dans mon travail : elle m'a permis de me concentrer sur la recherche, et de réaliser l'étude approfondie d'une langue sur la base de données collectées de première main. Les enquêtes de terrain confrontent le chercheur aux questions les plus variées : phonologie, morphosyntaxe et sémantique, dialectologie et sociolinguistique... Cela coïncide avec mon objectif d'une compréhension des phénomènes linguistiques en système et en contexte.

### 1.1.1. Etude de la langue na de Yongning

Après de courtes enquêtes aux villages naxi de Wenhua (visité en 2002), Guifeng et Fengke (visités en 2004), je me suis rendu en 2006 au village de Yongning, lieu plus excentré que la tradition linguistique chinoise inclut à l'intérieur du domaine de la langue « naxi » au sens large (Hé & Jiāng 1985). L'idée était d'obtenir une meilleure vue d'ensemble de la diversité dialectale du naxi et des langues qui sont ses plus proches parentes, en élargissant l'enquête à des dialectes qui n'ont pas été en contact régulier avec la langue naxi de la ville de Lijiang, centre administratif et politique de la région naxi depuis au moins sept siècles. Comme le montre la carte ci-dessous, la plaine de Yongning est située en haute altitude (au-dessus de 2 600 mètres ; la localité d'enquête est située à une altitude de 3 000 mètres), et est particulièrement enclavée.



Carte de Yongning, réalisée pour la monographie à paraître en 2017.

La population totale des locuteurs de la langue na de Yongning est estimée à 47 000 par les auteurs de la base de données *Ethnologue* (Lewis, Simons & Fennig 2016), mais ce chiffre approximatif recouvre une grande diversité de parlers dont certains n'ont à ma connaissance jamais été décrits. La langue na était dominante dans la plaine de Yongning jusqu'au milieu des années 1950 ; elle est actuellement en cours de remplacement par le chinois mandarin, de sorte que le travail de documentation linguistique présente un caractère d'urgence. J'ai travaillé essentiellement avec une locutrice née en 1950, dont les quatre enfants ont quitté Yongning pour raisons professionnelles et familiales, et sont plus à l'aise en mandarin qu'en chinois. Ses cinq petits-enfants, quoique tous élevés par elle (qui leur a toujours parlé en langue na au long de leur petite enfance), n'ont qu'une compréhension très imparfaite de la langue.

En parallèle avec le travail élémentaire qui consiste à établir l'inventaire phonémique et tonal, j'ai procédé à la recherche de régularités phonologiques qui régissent les alternances tonales. L'enquête a rapidement révélé que le système tonal du dialecte na du village d'Alawa, dans la plaine de Yongning – parler désigné ci-dessous comme *na de Yongning* – était d'une complexité sans commune mesure avec le naxi. Les tons en contexte (dans des phrases) sont souvent éloignés des réalisations tonales des mots pris isolément. De retour du premier séjour sur le terrain (en 2006), j'ai d'abord recherché des régularités d'ordre purement phonologique qui expliqueraient les formes de surface. Il existe en effet certaines règles phonologiques, telles que l'abaissement systématique (passage au ton bas) de tous les tons qui suivent un ton haut<sup>5</sup>. Mais celles-ci ne suffisent pas à expliquer l'ensemble des changements tonals : il s'est avéré que les règles d'ajustement tonal entre mots étaient *morpho*-tonologiques, et pas simplement phonologiques. Ces règles sont différentes selon le contexte morphosyntaxique. Ainsi, la combinaison de deux noms peut donner un résultat différent selon qu'ils se combinent en un composé déterminatif (tel que 'patte de poulet') ou en un composé coordinatif (tel que 'poule et poulet').

La rencontre avec ce système linguistique m'a fourni l'occasion de renouer progressivement avec la syntaxe, domaine de recherche abordé plus tôt dans mes études (le mémoire de recherche réalisé pour mon premier Diplôme d'Etudes Approfondies en Science du langage concernait l'objet du verbe en anglais britannique).

---

<sup>5</sup> Ces processus s'appliquent à l'intérieur d'une unité désignée comme le *groupe tonal* du fait que sa caractéristique définitoire est d'être le domaine d'application des règles tonales. Le chapitre 7 de l'ouvrage *Tone in Yongning Na : lexical tones and morphotonology*, joint au présent dossier, est consacré à cette unité phonologique, et aux liens qu'elle entretient avec la syntaxe, la sémantique et la stylistique.

Jusqu'en 2016, les enquêtes se sont succédées, aboutissant en 2016 à l'achèvement d'une monographie au sujet du système tonal du na de Yongning. L'ouvrage a été soumis le 1<sup>er</sup> juin 2016. C'est cette étape franchie qui m'a suggéré le projet de la présente candidature à l'Habilitation à Diriger les Recherches. Accepté le 21 octobre, le livre doit paraître au printemps 2017 (Michaud 2017)<sup>6</sup>. J'ai également préparé un dictionnaire de cette langue, et mis en ligne des enregistrements transcrits et annotés, afin de réaliser une documentation équilibrée, dans la tradition des linguistes dits « de terrain ».

One mainstay of the Boasian tradition in anthropological linguistics is the notion that adequate documentation of a language must consist of at least three volumes: a grammar, a dictionary, and a collection of texts. This convention grew out of Boas's dogged insistence on the collection of copious texts in the native languages as a way of documenting the cultures of Native North Americans, which he believed were breaking down and disappearing. Obviously, if one were actually to make use of such texts, a grammar and a dictionary were also needed; so this practice of a necessary trilogy was established, a tradition that has continued in academic departments which carry on the Boasian heritage (...). (Foley 1999:470)

Mon travail n'est pas typique de celui du linguiste de terrain dans la mesure où la monographie que j'ai écrite est centrée autour des tons, contrairement à ce qu'on serait en droit d'attendre d'une grammaire de référence équilibrée. Le choix de faire la part belle aux tons tient essentiellement à la richesse du système morphotonologique, qui appelle une description détaillée. Un second argument tient à l'existence d'une thèse intitulée *A descriptive grammar of Yongning Na* (Lidz 2010), grammaire généraliste d'un dialecte proche de celui que j'ai étudié : le na du village de Luoshui, sur les rives du lac Lugu (voir carte). L'existence de ce travail n'interdit certes pas le projet d'une nouvelle grammaire de référence, mais l'étude approfondie du système tonal (qui n'est pas central dans la thèse de L. Lidz) paraissait une priorité.

---

<sup>6</sup> Avant la parution de cette monographie, certains aspects du travail avaient fait l'objet d'une publication sous forme d'article ou de chapitre de livre : le système phonologique, y compris les tons lexicaux (Michaud 2008a) ; les propriétés acoustiques de certaines voyelles (Michaud & Vaissière 2009, faisant suite à une communication de colloque: Michaud & Vaissière 2008) ; les combinaisons entre numéraux et classificateurs nominaux (Michaud 2013a) ; les liens entre pragmatique et découpage de l'énoncé en groupes tonals (Michaud 2015a) ; et la problématique générale de la structure de l'information (Michaud & Brunelle 2016). Ce dernier travail, rédigé dans une perspective comparative, mérite quelques mots d'explication. J'ai été sollicité en 2011 pour la préparation d'un chapitre de livre au sujet de la structure de l'information dans les langues d'Asie, pour un manuel en préparation chez Oxford University Press : *Handbook of Information Structure*. Marc Brunelle a accepté de participer à ce travail. La demande des éditeurs était de présenter les problématiques liées à la structure de l'information dans les langues d'Asie, à l'exclusion du chinois, qui faisait l'objet d'un chapitre séparé. Comme le relevait avec humour un collègue, le domaine ainsi défini aurait inclus les langues ouraliennes, le youkaguir, le ket, l'iranien et le thaï dans un même chapitre. Même en se limitant aux seules langues sino-tibétaines, un rapide passage en revue de publications montre que la question de la structure de l'information se pose de façon très différente dans les trois cent et quelques langues du groupe. Marc Brunelle et moi avons choisi de présenter les langues dont nous sommes les plus familiers : le na et le vietnamien, dont la présentation selon un plan similaire fait ressortir les profondes différences au plan de la structuration de l'information.

Parmi les deux autres volets de la « trilogie boasienne », le *dictionnaire* et le *recueil de textes* bénéficient, via des archives différentes, des mêmes garanties de mise à disposition sur Internet et de pérennité dans le temps offertes par le système d'archivage pérenne mis en place en partenariat entre la Très Grande Infrastructure de Recherche Huma-Num, le Centre National d'Informatique de l'Enseignement Supérieur (CINES), et le Centre de calcul de l'Institut national de physique nucléaire et de physique des particules (In2P3). En effet, le dictionnaire est archivé via Hyper-Articles en Ligne – Sciences Humaines et sociales (HAL-SHS), et les textes via le réservoir de données Collection de Corpus Oraux Numériques (Cocoon), qui suivent *in fine* les mêmes procédures de versement à l'archive nationale. (Plus de détails au sujet de l'archive Cocoon sont fournis au §1.5.2.)

Sous l'impulsion de Guillaume Jacques, coordinateur d'un projet financé par l'Agence Nationale de la Recherche (projet HimalCo, ANR-12-CORP-0006, réalisé de 2013 à 2016), les données lexicales ont été mises en forme : conversion dans le format MDF (*Multi-Dictionary Formatter*, format natif des données dans le logiciel Toolbox, employé comme interface de saisie), et réalisation d'outils de conversion dans le format-pivot LMF (*Lexical Markup Framework*), lui-même employé pour la production de versions PDF mises en page (dictionnaire na-chinois-français et na-chinois-anglais) ainsi que d'une version électronique en ligne au format HTML. En 2015 a été mise en ligne la première version du dictionnaire na-chinois-français-anglais (Michaud 2015b). Le choix réalisé a été celui d'une édition purement électronique : deux fichiers PDF (na-chinois-français et na-chinois-anglais) sont mis en ligne et archivés de façon pérenne via HAL-SHS, accompagnés des fichiers-source (MDF et LMF) ; et le dictionnaire au format HTML est rendu disponible via l'interface de la Collection Pangloss. Le système de gestion des versions successives des documents mis en place au plan national pour l'archivage pérenne permet de publier commodément des versions améliorées. Afin de ne pas multiplier les versions, un rythme raisonnable serait de réaliser un nouveau dépôt à quelques années d'intervalle.

La réalisation d'un dictionnaire est une tâche de longue haleine. La version actuelle présente de nombreuses limites. D'une part, sa taille très restreinte : moins de 3 000 entrées ; par comparaison, le dictionnaire japhug de Guillaume Jacques (Jacques 2015) en compte 7 000. Le dictionnaire a vocation à être progressivement enrichi dans les années qui viennent, à mesure de la poursuite de mon travail. D'autre part, les deux documents PDF (na-chinois-français et na-chinois-anglais) dans leur première version (2015) présentaient quelques petits défauts techniques. Il a néanmoins paru préférable de mettre en ligne ces

documents .pdf en l'état dès 2015, accompagnés du fichier-source : en dépit de ses limites, le dictionnaire réalisé dans le cadre du projet HimalCo est incomparablement plus fiable que le seul autre dictionnaire de la langue na disponible auparavant (Zhībā & Xǔ 2013), qui comporte un tel nombre d'erreurs qu'il est à peu près inutilisable.

Parmi les trois volets de la « trilogie boasienne », c'est aux textes, base de tout le reste, que j'attache le plus de prix. Les enregistrements audio réalisés au fil de l'enquête sont progressivement mis en ligne depuis 2011 dans la Collection Pangloss, archive ouverte d'enregistrements audio (et également vidéo) du laboratoire Lacito, présentée en détail en section 1.5.2. Sont actuellement disponibles une vingtaine de récits transcrits et traduits, et des séances d'élicitation phonologique et morphotonologique, ainsi qu'une plus grande quantité de récits (plus d'une cinquantaine) qui n'ont pas encore été transcrits ni traduits. De même que pour les documents déposés dans l'archive HAL-SHS, les documents déposés dans l'archive Cocoon via la Collection Pangloss peuvent être améliorés au fil du temps, et de nouvelles versions déposées. L'utilisateur accède directement à la version la plus récente.

La possibilité de mettre à jour les transcriptions est d'une grande utilité, en particulier au cours des premières années où l'on apprend et décrit une langue, car les annotations produites par le linguiste reposent sur un ensemble d'analyses et d'hypothèses qui évoluent au fil du temps. Dans le cas du na de Yongning, mon analyse de certains points de phonologie segmentale, et mes choix de notation, ont quelque peu évolué depuis les premiers stades de l'enquête. Par exemple, j'ai d'abord cru remarqué que les voyelles [o] et [u] de mes premières notations phonétiques étaient en distribution complémentaire (une seule voyelle phonologique, deux timbres en fonction du contexte). J'ai hésité quant à la notation, choisissant dans un premier temps de noter le phonème comme /u/ de façon à employer des symboles vocaliques les plus éloignés les uns des autres dans le trapèze des voyelles, puis plus tard comme /o/, qui me paraît correspondre mieux à la réalisation phonétique majoritaire. Finalement, j'ai observé qu'il existait une opposition entre deux unités phonémiques après une initiale glottale, par exemple /ho/ dans 'perdrix' et /hu/ dans 'estomac', ce qui amène à reconnaître deux phonèmes, /u/ et /o/, dont l'opposition est neutralisée dans presque tous les contextes (Michaud 2017). De tels changements dans l'analyse nécessitent une mise à jour intégrale des documents déjà réalisés, afin d'éviter les confusions que créerait la coexistence, dans la collection de documents en langue na, de notations différentes pour une même unité phonologique.



A la date de rédaction du présent mémoire de synthèse, la notation phonologique est maintenant stabilisée, et les textes ont atteint un stade où les transcriptions et traductions sont assez fiables pour être réutilisables par d'autres que moi. Je compte poursuivre l'amélioration des textes et du dictionnaire au fil des années : je préfère approfondir cette langue au fil des années, et polir les documents recueillis avec un soin philologique, plutôt que d'ouvrir sur un autre terrain le même type de grand chantier qu'à Yongning. Laisser en plan un travail imparfait serait rendre un mauvais service aux collègues.

Dans ce travail d'établissement de textes et d'analyse des données, la comparaison entre langues intervient dans la mesure où des similarités notables sont décelées. N'ayant pas une culture linguistique suffisante pour identifier par moi-même, dans l'immense diversité des langues du monde, les termes de comparaison les plus heureux, j'emprunte avec gratitude les pistes signalées par des collègues, par exemple lorsque Guillaume Jacques attire l'attention des sinologues vers les langues mandé (Jacques 2014:6). Le parallèle paraît prometteur pour l'étude des langues naish (dont le na de Yongning), qui présentent des ressemblances avec les langues mandé sur le plan de la morphotonologie. Les comparaisons entre naish et mandé ébauchées dans la monographie consacrée au système tonal du na de Yongning (Michaud 2017) ont vocation à être poursuivies par la suite.

### **1.1.2. Etude d'autres parlars du groupe naish**

A la fin de l'année 2006, Ekaterina Chirkova, elle aussi fraîchement recrutée au CNRS, a proposé à Guillaume Jacques et moi-même de participer au montage d'un projet au sujet des langues du comté de Muli, voisin de Yongning, et où sont notamment parlés des dialectes proches du na et du naxi (parlars du groupe *naish* du sino-tibétain). Entamer un terrain de plus, alors que celui à Yongning commençait à peine, avait quelque chose d'intimidant ; mais d'un autre côté, ce projet coïncidait avec mon souhait d'explorer les parlars naish de cette région. En 2008 et 2009, dans le cadre du projet ANR PASQi (Phylogenetic Assessment of Southern Qiangic, ANR-07-JCJC-0063), j'ai mené deux enquêtes à Muli au sujet de la langue lazé, et collecté quelques données au sujet du parler na du village de Shuiluo.

Le lazé présente certaines spécificités intéressantes dans sa phonologie segmentale, par exemple une règle de dévoisement des occlusives voisées en position intervocalique dans les disyllabes, là où on attendrait l'inverse : renforcement du voisement des occlusives phonologiquement voisées, voisement des occlusives sourdes, et spirantisation des occlusives médiales. Son système tonal présente également quelques particularités notables : un passage au schéma H.H des disyllabes lors de leur lexicalisation, quels que soient les tons des deux monosyllabes qui les composent. Ces observations ont été présentées lors d'un Atelier à Taïwan en 2008<sup>7</sup>, puis publiées dans un article écrit en chinois (Michaud 2009). Les données lexicographiques recueillies ont été confiées à la base de donnée STEDT (en même temps que des données de langue na de Yongning) ; elles ont également été publiées en Annexe à une description phonologique publiée en Chine (Michaud & Jacques 2012). Deux récits avec gloses interlinéaires, ainsi que quelques documents audio non transcrits, sont en ligne dans la Collection Pangloss (au sujet de cette archive, voir §1.5.2).

D'autres travaux au sujet des langues naish ont été publiés au cours de la période 2006-2016, notamment des esquisses phonologiques de parlers naxi jusque-là non documentés : celui de Ciending (Michaud & Xu Jirong 2012), étudié avec un locuteur natif lorsque celui-ci était étudiant de Master (Université du Yunnan ; il n'a hélas pas poursuivi en thèse, mais envisage maintenant – 2017 – de reprendre des études doctorales), et celui de Pianding, étudié avec un locuteur natif qui est actuellement (2017) doctorant à Pékin (Michaud & He 2015) (voir §2.1.1 au sujet de nos projets communs).

Afin de faciliter les échanges avec les collègues chinois, et d'aider à la reconnaissance de mes travaux par les spécialistes du domaine en Chine, certains travaux ont été publiés en Chine (et en langue chinoise, à l'exception de deux publications en anglais). La distinction entre recherche « nationale » et « étrangère » est en effet particulièrement nette en Chine. Certains travaux ont été rédigés directement en chinois (Michaud 2009; Michaud & Xu Jirong 2012), d'autres ont été traduits par des collègues chinois et revus par mes soins (Michaud 2013b; Michaud, Jacques & Rankin 2014). J'ai également traduit un article de 2006 au sujet du système de transcription de la langue naxi employé par un pionnier de l'étude des Naxi : le botaniste et explorateur Joseph Rock (1884-1962), qui séjourna parmi les Naxi à plusieurs reprises entre 1922 et 1949 (Michailovsky, Michaud & He Xueguang 2015).

---

<sup>7</sup> Le texte intégral en est disponible ici : <https://halshs.archives-ouvertes.fr/halshs-01372956/>

Ces tâches de traduction sont gourmandes en temps. Les textes que j'écris en chinois doivent être revus par un locuteur natif, après quoi je vérifie en détail les modifications suggérées, réfléchissant au choix des termes et veillant à ce que leur emploi ne soit pas inconséquent. Les procédures éditoriales sont également susceptibles d'être longues et complexes. A titre d'exemple, la traduction d'un article au sujet des notations phonétiques de Joseph Rock, publiée en 2015 (Michailovsky, Michaud & He Xueguang 2015), a ses origines dans une traduction réalisée quatorze ans auparavant. En 2001, à l'occasion d'une visite à Paris, un collègue chinois avait encouragé Boyd Michailovsky à publier en Chine un de ses travaux en traduction chinoise. Après mûre réflexion, Boyd Michailovsky avait choisi un manuscrit inédit dans lequel il expliquait comment il avait établi des correspondances entre ses notations du naxi en alphabet phonétique et l'alphabet idiosyncratique de l'explorateur-botaniste Joseph Rock (1884-1962). Mesurant l'utilité de cet article pour des lecteurs chinois qui avaient accès au texte de Rock mais ne savaient pas interpréter ses notations, je l'avais traduit avec grand soin, à la demande de B. Michailovsky. Par la suite, celui-ci était resté sans nouvelles de l'éditeur. Enfin, il lui avait été signifié que la revue ne pouvait publier ce texte, et qu'il pourrait tout au plus figurer temporairement sur son site internet. Renseignements pris, il s'est avéré que l'article avait paru trop peu riche en contenu « théorique ». Au nom d'une hiérarchie entre théorie (l'aboutissement du travail de recherche) et matériaux de recherche (la documentation : matière première à faible valeur ajoutée), les éditeurs s'étaient formalisés que Boyd leur propose un article de nature documentaire, et partant, « secondaire », plutôt qu'un article de « recherche théorique ».

Boyd Michailovsky avait su prendre l'épisode avec humour, sans se formaliser. Par la suite, nous avons développé ensemble l'article, sur la base de données nouvelles collectées de 2002 à 2004. L'article en anglais préparé en collaboration était paru en 2006 (Michailovsky & Michaud 2006).

Le manuscrit chinois a ensuite été proposé au traducteur chinois du dictionnaire de Rock, comme un outil de correspondance à faire figurer en annexe au volume qu'il était en train de traduire ; mais l'auteur n'a pas donné suite. Après coup, j'ai compris qu'il avait été retenu par certaines craintes au sujet de la fiabilité de cet outil : n'étant pas linguiste, il n'était pas familier de la méthode par laquelle les correspondances avaient été établies (comparaison systématique de prononciations des mots dans la langue orale avec les transcriptions du dictionnaire), de sorte qu'il était en peine de se déterminer sur leur valeur.

Faire figurer un outil élaboré par d'autres chercheurs dans le volume qu'il traduisait lui paraissait en outre problématique.

L'outil de correspondance entre notations de Rock et alphabet phonétique commençait néanmoins à circuler, dans sa version anglaise, dans le petit cercle des lecteurs intéressés. Des collègues de Lijiang (région où est parlée la langue naxi) m'ont encouragé à réaliser une nouvelle version chinoise et à la publier. J'ai donc soumis une nouvelle traduction à une revue publiée dans la province du Yunnan, *Chama Gudao Yanjiu Jikan* 茶马古道研究集刊, dont le premier numéro était en cours d'élaboration. J'ai simultanément soumis à la même revue un autre travail traduit, au sujet de l'analyse phonologique de la langue naxi. C'est ce dernier qui a été publié en premier (Michaud & He Xueguang 2010), tandis que la traduction du travail en collaboration avec Boyd Michailovsky était à nouveau reportée, les éditeurs de cette revue considérant eux aussi que le contenu était trop peu riche en théorie. J'ai repris contact avec les éditeurs en 2012, puis en 2013. Ils m'ont informé qu'ils avaient un temps envisagé de publier cet article dans un recueil (ouvrage collectif), mais que ce projet de volume n'avait pas fait grand progrès et qu'ils envisageaient maintenant une publication en 2014. La traduction a finalement été publiée en 2015 (Michailovsky, Michaud & He Xueguang 2015).

En dépit des délais, le partenariat avec cette revue s'est avéré fructueux. En 2012, j'ai été nommé Membre associé du Centre de recherches qui édite cette revue : le Centre de recherche sur la culture de la Route du Thé 云南大学茶马古道文化研究所 (Université du Yunnan, Kunming, Chine). Cette revue a successivement publié cinq de mes travaux en chinois, certains originaux, d'autres adaptés de publications anglaises. La revue est peu diffusée, et n'a pas actuellement de version électronique ; mais dès qu'un article est accepté pour publication, je dépose dans l'archive HAL la version pré-publiée, qui est dès lors archivée de façon pérenne, et devient accessible aux internautes. Les textes téléchargés via HAL peuvent être cités avec la référence de publication chinoise. L'accès aux sites étrangers depuis la Chine est soumis à des aléas contrôlés avec doigté, de sorte que les documents hébergés par HAL ne sont pas nécessairement accessibles pour les internautes chinois, mais les lecteurs motivés et patients peuvent en règle générale avoir accès aux documents.

L'exemple de la traduction réalisée en 2002, remaniée en 2006 et publiée en 2015 est quelque peu extrême, mais l'investissement de temps que requiert la publication d'articles en chinois est bien réel. Cet investissement ne paraît pas absurde au vu de l'enjeu :

l'accès de collègues en Chine aux résultats de mes recherches, publiés en chinois. J'ai publié deux travaux dans une autre revue, elle aussi peu diffusée, publiée localement dans la ville de Lijiang, centre de la région où est parlé le naxi. L'un expose l'objectif et les méthodes de la documentation linguistique au sujet de la langue na de Yongning (Michaud 2012a). L'autre (Michaud & Lātāmī 2010:20) constitue une traduction d'un travail au sujet des « oppositions phonologiques en danger » dans ce parler. La version anglaise de ce texte est parue chez un éditeur chinois, avec un co-auteur chinois (Michaud & Latami Dashi 2011), ce qui contribue aussi à rendre mes travaux moins lointains pour les lecteurs chinois.

Ces publications en chinois ont sans doute joué un rôle dans la décision prise par une deuxième institution (en 2012) de m'accorder le statut de Membre associé : l'Institut de recherche sur la culture dongba (丽江市东巴文化研究院, Lijiang, Chine). La reconnaissance de mon travail par certaines institutions chinoises permet de bénéficier de leur soutien à l'occasion des demandes de visas et d'autorisations de recherche pour accéder au terrain en Chine. Les autorisations demeurent difficiles à obtenir, même avec l'appui d'institutions locales. Le pays ne soutient pas particulièrement l'implication de chercheurs étrangers dans des recherches concernant les « minorités ethniques », qui constituent un sujet sensible, lié aux questions d'identité nationale. Les collègues chinois sont les mieux à même de suggérer des moyens – sans cesse à réinventer – pour que le travail puisse continuer malgré un contexte parfois difficile.

## 1.2. Phonétique expérimentale et phonologie théorique

Documents joints au dossier concernant ce volet : 2 articles :

Michaud, Alexis & Xueguang He. 2007. Reassociated tones and coalescent syllables in Naxi (Tibeto-Burman). *Journal of the International Phonetic Association* 37(3). 237–255.

Mazaudon, Martine & Alexis Michaud. 2008. Tonal contrasts and initial consonants: a case study of Tamang, a “missing link” in tonogenesis. *Phonetica* 65(4). 231–256.

### 1.2.1. Phonétique expérimentale

La phonétique expérimentale constitue mon domaine central de spécialisation et d'expertise reconnue. Mon rôle de « phonéticien de terrain », tel que je le conçois, consiste à réaliser un *gros plan* expérimental sur tel ou tel point précis identifié au fil de l'étude d'un système phonologique : « chaînon manquant » dans l'évolution diachronique, ou phénomène contrevenant aux généralisations typologiques qui paraissent les mieux établies. A titre d'exemple, les schèmes de reduplication du naxi ont de quoi laisser perplexe un phonologue, ce qui encourageait à formuler des hypothèses originales au sujet de leur origine. L'analyse menée en collaboration avec Jacqueline Vaissière (Michaud & Vaissière 2007a; version anglaise: Michaud & Vaissière 2007b) souligne la proximité phonétique qui existe entre les formes redupliquées avec changement catégoriel de ton (type AABB) et sans changement catégoriel (type ABAB). Cette observation suggère que les deux ensembles de formes redupliquées pourraient avoir pour origine commune une reduplication à l'identique. L'évolution ultérieure de deux des schémas serait due au figement de tendances phonétiques dont l'effet demeure observable en synchronie. La variation phonétique de certains tons aurait été réinterprétée comme une différence phonologique de catégorie tonale. Une réflexion est proposée au sujet des conditions historiques qui auraient pu permettre cette phonologisation. Ainsi, l'analyse conjugue observations phonologiques, expérimentation phonétique, et hypothèses diachroniques.

Ainsi entendue, ma spécialisation en phonétique/phonologie n'est pas si éloignée des pratiques les plus courantes des linguistes du Lacito et des laboratoires thématiquement les plus proches (LLACAN, SEDYL, CRLAO...). Cette spécialisation (qui n'est nullement exclusive d'un intérêt pour d'autres domaines de la linguistique) s'est avérée être source de projets communs avec des collègues non phonéticiens qui ont le souci d'approfondir des

questions de phonétique/phonologie rencontrées au cours de leurs travaux au sujet des langues les plus diverses.

Depuis ma thèse, j'ai publié deux articles dans des revues de phonétique : une étude réalisée pendant ma thèse sur un point de tonologie naxi (la place des contours montants dans le système) est parue en 2007 dans *Journal of the International Phonetic Association* ; et un travail en collaboration avec Martine Mazaudon, spécialiste de la langue tamang, est paru en 2008 dans *Phonetica*.

### **1.2.1.1. Une étude au sujet de la réduction syllabique en langue naxi**

L'article dans dans *Journal of the International Phonetic Association* au sujet du naxi (Michaud & He Xueguang 2007) fait suite à une étude non expérimentale parue dans *Linguistics of the Tibeto-Burman Area* (Michaud 2006a)<sup>8</sup>. Il présente un ensemble d'expériences qui établit que les séquences tonales Moyen + Haut et Bas + Haut (sur une seule syllabe), consécutives à la réduction d'une syllabe au ton Haut, ont des réalisations distinctes (non homophones), contrairement à ce que laissait entendre la notation qui était employée jusque-là, laquelle ne reconnaissait qu'un unique contour montant (Hé & Jiāng 1985). Sur la base de l'analogie avec des faits semblables dans des langues subsahariennes, il paraît à la fois adéquat et économique de parler de « ton Haut flottant » au sujet du naxi, cela d'autant plus que l'ajout d'un tel ton Haut produit des contours différents selon qu'il s'ajoute à un ton Bas ou Moyen (contour Bas-Haut et Moyen-Haut, respectivement, conformément à ce qui est attendu pour un ton flottant). La réduction de syllabes au ton Haut présente un caractère catégoriel, et une nette similarité avec les « tons flottants » attestés dans d'autres langues (tels qu'ils sont par exemple décrits par Clements *et al.* 1984). La tradition des études orientalistes est plutôt orientée vers la réflexion diachronique, d'où des formulations qui s'éloignent parfois sensiblement de ce à quoi aboutirait une analyse strictement synchronique, nourries qu'elles sont de connaissances concernant des états plus anciens des langues, jusqu'aux états où celles-ci ne possédaient pas encore de tons, et avaient en revanche une structure syllabique plus complexe au plan consonantique. Il est donc compréhensible que la notion de ton flottant soit née, non dans les études orientalistes (où a été proposée l'expression de « meaningful tone » : Ratliff 1992), mais dans les études

---

<sup>8</sup> Ces deux articles font partie du dossier de publications de la présente candidature à l'Habilitation.

africanistes (en particulier dans l'étude des langues Niger-Congo), où ces processus sont très productifs.

Larry Hyman (spécialiste des langues Niger-Congo, en particulier) analyse les tons de certaines langues tibéto-birmanes, les langues kuki, en termes de tons Bas et Hauts, et y décèle des processus « africains » (l'expression est de lui; elle est également employée par Evans 2008), tels que le *downstep* (Hyman & VanBik 2002; Hyman & VanBik 2004; Hyman 2010). René Gsell suggérait également une analyse de ce type pour la langue lushai (mizo) (Gsell 1979:8). Ces découvertes appellent approfondissement, dans un dialogue entre traditions linguistiques. L'examen détaillé de langues dans lesquelles les traits « flottants » ont une place restreinte (quoique bien réelle) doit permettre de préciser dans quelle mesure les faits sont identiques, au plan phonétique et au plan morphophonologique. Tel est un des enjeux de la monographie au sujet des tons du na de Yongning (Michaud 2017), pièce maîtresse du présent dossier, qui parle – je l'espère – pour elle-même, et que je n'essaierai pas de résumer ici.

L'article de 2007 au sujet du naxi a attiré l'attention d'un étudiant de linguistique, Likun He, locuteur natif de cette langue, qui envisage de reprendre et approfondir ce thème de recherche, dont je n'ai pas poursuivi l'exploration parce que je m'employais à explorer d'autres parlars du groupe naish. (Des projets communs avec Likun He sont évoqués dans la section « Orientations futures ».)

### **1.2.1.2. Etude de la langue tamang**

Une collaboration avec Martine Mazaudon avait été entamée en 2004, à son initiative, avec pour objet l'étude approfondie de la réalisation phonétique des quatre tons de cette langue sino-tibétaine parlée au Népal. Ces tons sont d'un grand intérêt pour la typologie prosodique, du fait que leur domaine est le mot phonologique, non la syllabe, et que leur réalisation est très variable, les locuteurs faisant flèche de tout bois dans la réalisation des contrastes tonals : hauteur/longueur/intensité, qualité de voix, et articulation des consonnes. Tout bois, ou presque : le timbre vocalique, quoique non contraint par la taille – restreinte – de l'inventaire vocalique, ne paraît pas fournir d'indices sur la catégorie tonale du mot tamang. Ce fait est difficilement compréhensible dans le cadre de l'hypothèse strictement synchronique d'un partage des ressources entre spécifications lexicales et intonation (« tout ce qui n'est pas employé à un niveau est disponible pour les autres niveaux »). Si les nuances fines de timbre ne sont pas distinctives pour opposer des



catégories vocaliques, on s'attendrait à ce qu'elles soient utilisées comme corrélats secondaires d'autres oppositions (tel est le cas en môn, où l'opposition lexicale entre syllabes en voix claire et syllabes en voix soufflée en môn a une incidence très nette sur le timbre vocalique). En tamang, ce fait paraît justiciable d'une explication diachronique, les tons restant liés aux consonnes (dont ils sont issus) plus qu'aux voyelles.

Mon apport a consisté à mettre en place un protocole expérimental susceptible de confirmer aux yeux d'un public de phonéticiens/phonologues les observations réalisées par Martine Mazaudon (Mazaudon 1973). Il s'agissait en particulier de quantifier les observations au sujet de la « voix soufflée » associée à deux des quatre tons du système tamang. Sans s'arrêter au constat de variabilité, qui constitue un *invariant* de la phonétique expérimentale contemporaine, l'orientation de ces recherches consiste dans un premier temps à rechercher quel paramètre présente le moins de variabilité, et à modéliser la stratégie d'un locuteur. Dans un second temps, les différentes stratégies individuelles sont comparées pour faire ressortir un fonds commun.

Le travail expérimental entamé en 2004 sur la base d'enregistrements audio anciens s'est ensuite poursuivi au moyen de nouvelles données audio et électroglottographiques. Cette étude permet tout à la fois de dresser un tableau synchronique de ces tons atypiques et de prendre la mesure des changements intervenus au cours des trente années passées. D'après la description proposée par Martine Mazaudon au sujet de la langue telle qu'elle était parlée au début des années 1970, les tons 3 et 4 étaient réalisés avec une phonation soufflée ; dans les données enregistrées en 2005 auprès de cinq locuteurs âgés de 30 à 45 ans, le ton 4 est moins nettement soufflé que le ton 3, et deux des cinq locuteurs paraissent ne faire usage d'une voix soufflée sur aucun des quatre tons.

La démonstration expérimentale, indispensable à la reconnaissance du travail par les collègues phonéticiens/phonologues, est établie (Mazaudon & Michaud 2005; Michaud & Mazaudon 2006; Mazaudon & Michaud 2008) ; la quantification des résultats a permis d'affiner le tableau synchronique, et de mettre en lumière une différence entre la génération de locuteurs représentée dans l'expérience (trentenaires enregistrés en 2005) et celle de leurs grands-parents (personnes âgées, enregistrées au début des années 1970). Martine Mazaudon poursuit actuellement (en 2016-2017) l'exploration de ces questions, pour d'autres parlers du même groupe, en collaboration avec Jiayin Gao, dans le cadre d'un post-doctorat au Lacito. (Un billet des Carnets du Lacito fournit plus d'informations : <http://lacito.hypotheses.org/332>.)

### 1.2.1.3. Etude expérimentale de la nasalité

Mon premier travail de phonétique expérimentale, au sujet de certaines oppositions syllabiques en vietnamien de Hanoi (Michaud 2004), m'avait amené à examiner la réalisation de la nasalité dans des rimes à nasale finale /m/, /n/, /ŋ/ dans les cas où le ton, glottalisé, interrompt brutalement la syllabe, avant qu'un segment nasal ne soit réalisé. Une question qui se posait était : dans quelle mesure la voyelle qui précède est-elle nasalisée, assurant ainsi la distinction entre des rimes comme /am/, /an/, /aŋ/ d'une part, /ap/, /at/, /ak/ d'autre part. L'observation spectrographique suggérait paradoxalement qu'une nasalisation était présente dans les syllabes à occlusive finale, aussi bien (voire plus encore) que dans les syllabes qui au plan phonémique contiennent une consonne nasale finale. Pour poursuivre l'étude de cette question, des données de débit d'air nasal et oral ont été prises. Une étude-pilote avec des données d'un locuteur a été réalisée et publiée (Michaud et al. 2006)<sup>9</sup>, mais mes autres projets ne m'ont pas permis de poursuivre l'étude de ce point de phonétique/phonologie du vietnamien. Les faits sont assez originaux, et les premiers résultats assez encourageants, pour suggérer le projet d'un Master voire d'une thèse en phonétique expérimentale (comprenant un volet perceptif). Cela fait partie des sujets que je pourrais proposer à des étudiants intéressés par la nasalité ou par la phonétique du vietnamien de Hanoi.

## 1.2.2. Phonologie théorique

A l'invitation de Nick Clements, j'ai participé à une réflexion commune avec lui et Cédric Patin au sujet des traits tonals, de 2004 jusqu'à la disparition de Nick en août 2009. Ce travail publié en 2011 (Clements, Michaud & Patin 2011) a été essentiellement planifié et réalisé par Nick Clements. La question de recherche initiale, en elle-même – existe-t-il des traits tonals universels ? –, se fonde sur le modèle de la grammaire générative : proposer un modèle qui permette de décrire toutes les langues. Nick Clements savait que je ne partageais pas ses convictions chomskyennes, mais cela ne l'avait pas empêché de m'encourager, et d'être toujours disponible pour des discussions, au sujet des tons en particulier, au cours desquelles il me faisait bénéficier de son impressionnante connaissance

---

<sup>9</sup> La collaboration avec Angélique Amelot entamée à l'occasion de cette étude de la nasalité en vietnamien s'est étendue à une étude de données françaises qu'elle avait enregistrées (Amelot & Michaud 2006).

de la bibliographie phonologique. Il me serait donc paru mal venu de me défilier lorsqu'il nous a proposé, à Cédric Patin et à moi, de collaborer à un article au sujet des traits tonals.

Les conclusions du chapitre peuvent paraître minces, comme le relève Robert Ladd dans une recension du volume : la recherche de *traits universels* est en décalage avec certaines discussions contemporaines.

...the logic (...) amounts to this: segmental features are universal; tone features seem not to be universal; therefore tones are different from segments. They do not question the major premise that segmental features are universal. Odden's paper is different. Right from the start he explicitly argues that 'the basic source of the problem of answering these questions [about tone features] lies in incorrect assumptions about the nature of features, specifically the assumption that there is a single set of predetermined features with a tight, universal mapping to phonetics' (81). (...) On the whole I agree with Odden that the conventional understanding of features is shot through with contradictions, which I discuss at length in the chapter on features and autosegments in Ladd 2014. (Ladd 2013)

Une seconde limite de ce travail est que Nick est demeuré convaincu qu'une analyse plus ingénieuse que celles tentées jusqu'à présent pourrait révéler la possibilité d'une analyse en niveaux (tons ponctuels) même pour les systèmes en apparence les plus réfractaires à ce type d'analyse, par exemple le vietnamien ou les langues sinitiques. Cinq années d'échanges n'ont guère fait évoluer les positions des coauteurs. Pour ma part, familier aussi bien de systèmes à tons ponctuels, par exemple en naxi, lazé et na, que de systèmes à tons complexes, comme le vietnamien ou le mandarin, je suis convaincu que la typologie linguistique doit reconnaître divers modes d'organisation tonale. Dans le travail en commun avec Nick, ces différences de points de vue se traduisent par des passages comme le suivant : « (The issue whether all tone systems can be analyzed in terms of levels and scales is left open here.) » Par certains côtés, le chapitre est donc un compromis entre points de vue que les auteurs ne sont pas parvenus à réconcilier.

Le chapitre n'en possède pas moins certaines qualités. La réanalyse par Nick Clements des faits de langues africaines, yala, bariba et surtout ewe, constitue une avancée par rapport aux analyses qu'il avait précédemment publiées. Ces analyses, abondamment citées à l'appui de modèles postulant un trait de registre (synchronique) pour les langues à quatre tons, s'avéraient erronées. En outre, dans cet article de phonologie théorique il est fait appel à une perspective diachronique pour éclairer les faits, ce qui est tout à fait en phase avec les méthodes qui me paraissent les plus prometteuses.

## 1.3. Prosodie : études de cas et réflexions générales

### 1.3.1. Etudes de cas

Dans le domaine de la prosodie, un travail en collaboration avec Katia Chirkova (CRLAO-CNRS) au sujet du système prosodique de la langue shixing (aussi appelée xumi ; famille sino-tibétaine, Chine) a été planifié en 2007 et réalisé au cours de l'année 2008. Le shixing est une langue qui présente des similarités frappantes avec le naxi. L'étude du système prosodique de plusieurs langues à la fois apparentées (au plan diachronique) et relativement proches au plan synchronique s'avère fructueuse : cela permet de disposer d'emblée d'une « trousse à outils » relativement bien adaptée, des techniques d'élicitation jusqu'aux outils pour la modélisation. Le travail au sujet du shixing a débouché sur un article de revue (Chirkova & Michaud 2009). Pour le premier auteur (E. Chirkova), il a fourni l'occasion d'un premier contact avec la phonétique expérimentale, discipline pour lequel son intérêt s'est confirmé par la suite, comme en témoignent ses publications (par exemple Chirkova & Chen 2013a; Chirkova & Chen 2013b; Chirkova, Chen & Antolík 2013; Chirkova et al. 2015).

J'ai également accédé à la demande de Mme Denise Bernot (fraîchement octogénaire à l'époque) qui souhaitait tirer le portrait aux tons du birman, au moyen des meilleurs outils de la phonétique expérimentale. Des enregistrements audio et électroglottographiques ont été réalisés, et des tracés en ont été extraits, qui lui ont fourni la matière d'exposés.

J'ai en outre participé à des travaux concernant le français (Vaissière & Michaud 2006) et le vietnamien (Michaud 2004; Michaud et al. 2006; Vu-Ngoc Tuân, d'Alessandro & Michaud 2005; Nguyen et al. 2013; Mac et al. 2015), et continué à publier des résultats obtenus dans ma thèse au sujet du naxi (notamment une caractérisation phonétique des trois tons ponctuels du naxi: Michaud, Vaissière & Nguyễn 2015). Un travail commun avec Vu-Ngoc Tuân (Michaud & Vu-Ngoc 2004) a été publié en traduction vietnamienne (Michaud & Vu-Ngoc 2007).

### 1.3.2. Réflexions générales et typologie prosodique

Des réflexions générales au sujet de la typologie prosodique (tons et intonation) étaient proposées dans ma thèse, sur la base d'une comparaison entre données anglaises, vietnamiennes et naxi. Le résumé court de la thèse était le suivant :

La mise en regard de données d'anglais et de deux langues à tons d'Asie (naxi et vietnamien) montre un *partage des ressources* entre spécifications lexicales (oppositions phonémiques, et accentuation) et intonation : les premières fournissent le cadre dans lequel la variation intonative prend son sens. Au plan typologique, chaque langue offrirait à l'énonciateur un certain degré de liberté ; le naxi (4 tons) apparaît plus contraint que l'anglais, et que le vietnamien (qui pourtant possède 6 tons). Les expériences menées (qui comportent une évaluation de la *qualité de voix* par électroglottographie) paraissent faire ressortir la *superposition* de plusieurs phénomènes. Elles permettent d'aborder une question soulevée par le courant *autosegmental-métrique* d'études intonatives : accentuation et intonation concourent-elles à une unique *séquence tonale* pour l'énoncé (cela dans toutes les langues) ? Il semblerait que ce dernier point de vue ne permette pas de rendre compte des faits rencontrés.

Mon approche est influencée par l'enseignement de Jacqueline Vaissière (Vaissière 2002; Vaissière 2004), qui reconnaît dans la prosodie la superposition et l'interaction de plusieurs niveaux : essentiellement, niveau lexical (tons ou accent, selon la langue), niveau « syntaxique » (découpage en groupes), et niveau « pragmatique » (reflétant la structure de l'information). Cette approche pourrait être classée comme « superpositionniste », dans la mesure où elle se distingue des modèles linéaires dits « autosegmentaux » qui s'imposent un inventaire très réduit de primitives.

Une partie des résultats expérimentaux de ma thèse ont été exposés au colloque *Speech Prosody 2006*, sous le titre « Replicating in Naxi (Tibeto-Burman) an experiment designed for Yorùbá: an approach to 'prominence-sensitive prosody' vs. 'calculated prosody' » (Michaud 2006b). Cette communication ébauchait une comparaison entre naxi et yoruba, langues qui toutes deux possèdent trois tons de base (Haut, Moyen et Bas) mais présentent des différences instructives dans la réalisation d'énoncés dont toutes les syllabes portent le même ton. En comparaison de langues subsahariennes telles que le ngamambo, le yoruba, l'igbo ou le mambila, les langues d'Asie que sont le naxi, le vietnamien et le chinois mandarin présentent des similarités entre elles. L'étude de ces questions paraît nécessiter l'emploi de protocoles similaires de langue à langue, et d'un vocabulaire descriptif homogène. C'est par une analyse détaillée d'exemples précis qu'il est possible d'avancer.

A titre d'exemple, un travail réalisé avec Barbara Kühnert (Michaud & Kühnert 2006; version anglaise: Michaud & Kühnert à paraître) examine un point qui fait débat dans les études prosodiques. Certains auteurs divisent la syllabe en initiale et rime : ainsi, dans une syllabe /man/, l'initiale est /m/ et la rime /an/ ; le ton appartient à la syllabe, mais est porté par la rime (voir par exemple Sagart 1993:35). Les descriptions de diverses langues font écho à cette division en initiale et rime (au sujet du yorùbá, langue de la famille niger-congo, voir Laniran 1992:61; au sujet du danois: Gårding 1998:127). Mais diverses études, portant également sur des langues variées, considèrent au contraire que la courbe de  $F_0$  portée par les consonnes initiales sonantes est partie intégrante de la courbe de  $F_0$  qui caractérise la syllabe dans son entier (en anglais: Ladd et al. 1999; en chinois: Xu & Wang 2001; en grec: Arvaniti, Ladd & Mennen 1998; en espagnol mexicain: Prieto, van Santen & Hirschberg 1995; en catalan: Prieto 2009; en arabe: Chahal & Hellmuth 2014). Ces études reposent sur l'idée qu'à un certain niveau d'abstraction, la hauteur serait une ligne continue, en dépit du non-voisement de certains segments. Pour progresser dans l'analyse de cette question, l'angle expérimental adopté a consisté à comparer des syllabes à initiale aspirée, non aspirée et nasale dans un même contexte prosodique (phrases lues). L'expérience-pilote nous conduit à conclure que dans le cas des monosyllabes anglais étudiés, la portion de la courbe de  $F_0$  portée par la consonne initiale doit être écartée pour que ressorte, dans son unité, le phénomène prosodique qui se réalise sur la syllabe étudiée, lequel entre dans la catégorie *Fall* de la tradition britannique d'études intonatives.

Mais au cours des années qui ont suivi, engagé dans l'exploration de langues parlées dans le Yunnan (Chine), je n'ai pas poursuivi dans le droit fil de tels travaux expérimentaux. En particulier, je n'ai pas continué la comparaison expérimentale détaillée entre la prosodie de l'anglais et celle de langues tonales (naxi et vietnamien). Or les propositions concernant la prosodie se prêtent à une confrontation aux données par une implémentation logicielle en synthèse de la parole – technique dite de l'analyse par synthèse (Öhman & Lindqvist 1968)). Elles peuvent également être explorées au moyen de techniques avancées d'observation de la production de parole (Post et al. 2015), de l'acoustique (Ferragne & Pellegrino 2004; Rouas et al. 2005; Ferragne & Pellegrino 2008), et de la perception. N'étant guère actif dans ce domaine expérimental au cours de la période concernée (2006-2016), j'ai choisi de rester en retrait des discussions générales dans ce domaine.

J'ai néanmoins eu l'occasion de formuler mon point de vue à la sollicitation de collègues, sous forme d'exposés invités (*keynote speeches*) au Séminaire international « Production de la parole » ISSP 2008 (Michaud 2008b), au symposium TAL 2012 (« Tonal Aspects of Language 2012 » : Michaud 2012b) et au Colloque international au sujet de la phonétique des langues de Chine ICPLC 2013 (Michaud 2013c). Suite à l'exposé à TAL 2012, ces réflexions ont été développées en collaboration avec Jacqueline Vaissière dans un article publié dans un recueil de *Working papers* (Michaud & Vaissière 2015). Ce type d'exposé amène à relever les divergences de vues avec certains modèles actuellement dominants, ce qui, d'après mon expérience, a vite fait de susciter des tensions, quelque effort que l'on déploie pour tenter de désamorcer la polémique. Comme le remarquait un collègue, il paraît plus sage d'attendre que les collègues réalisent par eux-mêmes que les notations telles que ToBI (Silverman et al. 1992) ne rendent pas les services attendus, pour proposer alors d'autres solutions, plutôt que d'aller à l'affrontement en critiquant des modèles qui gardent actuellement un vif attrait dans la communauté des chercheurs en prosodie. Il existe actuellement des initiatives pour élaborer de notations syncrétiques, qui rendent interoperables des corpus annotés avec des outils qui diffèrent profondément entre eux : par exemple DIMA (pour Deutsche Intonation, Modellierung und Annotation: Kügler et al. 2015). Cela peut aboutir à enrichir progressivement les modèles intonatifs, et au final, à refermer sans bruit la parenthèse des modèles binaires de l'intonation.

Dans l'immédiat, je m'abstiens de rompre inutilement des lances au sujet de l'intonation et de la typologie prosodique. Il existe un risque de « perdre le temps précieux à la recherche sur la chose même en se maintenant dans le choc des discours »<sup>10</sup>. Pour autant, je ne décline pas les sollicitations, par exemple pour la rédaction d'un article de synthèse « Prosodic systems of Southeast Asia » (avec Marc Brunelle – premier auteur –, James Kirby et Justin Watkins) pour un manuel dirigé par C. Gussenhoven et Aoju Chen : *Handbook of Prosody* (Oxford University Press, prévu pour 2018). On trouverait aisément des sujets de désaccord dans le projet d'introduction que les éditeurs ont fourni aux contributeurs, par exemple leur choix du cadre dit « autosegmental-métrique » dont ils relèvent (comme une de ses qualités) que « this mainstream phonetics-phonology model effectively merged the study of intonation and the study of tone », fusion que je serais porté à considérer comme un grand pas en arrière pour les études prosodiques. Mais le fait

---

<sup>10</sup> Formule empruntée à Paul-Laurent Assoun, *Psychoanalyse*, Paris : Presses Universitaires de France, 2007, p. XII.

que les éditeurs du volume aient souhaité que je participe à l'équipe des auteurs (et que ceux-ci ne soient pas contrariés de me voir à leurs côtés) paraît un signe encourageant de leur acceptation d'une certaine diversité des points de vue. L'invitation des éditeurs était assortie du message selon lequel les auteurs auraient toute liberté de ne pas s'en tenir au cadre théorique précis annoncé pour l'ensemble du volume. Cette situation, dans laquelle des malentendus peuvent naître à chaque détour d'une phrase, par l'emploi de termes que chacun comprend dans un sens différent, a quelque chose de malcommode ; mais, faisant de nécessité vertu, on relèvera qu'un tel contexte aide à se concentrer sur l'essentiel, et à l'expliquer sans ambages, dans un effort constant pour atteindre à la plus grande clarté.

## 1.4. Phonétique historique

Documents joints au dossier concernant ce volet : 4 articles :

- Jacques, Guillaume & Alexis Michaud. 2011. Approaching the historical phonology of three highly eroded Sino-Tibetan languages: Naxi, Na and Laze. *Diachronica* 28(4). 468–498.
- Michaud, Alexis. 2011. The tones of numerals and numeral-plus-classifier phrases: on structural similarities between Naxi, Na and Laze. *Linguistics of the Tibeto-Burman Area* 34(1). 1–26.
- Michaud, Alexis, Guillaume Jacques & Robert L. Rankin. 2012. Historical transfer of nasality between consonantal onset and vowel: from C to V or from V to C? *Diachronica* 29(2). 201–230.
- Michaud, Alexis, Michel Ferlus & Minh-Châu Nguyễn. 2015. Strata of standardization: the Phong Nha dialect of Vietnamese (Quảng Bình Province) in historical perspective. *Linguistics of the Tibeto-Burman Area* 38(1). 124–162.

Je partage avec nombre de collègues la conviction selon laquelle la linguistique historique est d'une importance capitale pour la bonne compréhension des faits linguistiques. En particulier, la phonologie diachronique est une composante qui devrait être centrale dans les études de phonétique/phonologie. Peu expérimenté dans ce domaine, je préfère m'y aventurer en bonne compagnie, en faisant équipe avec des spécialistes.

Je suis à même d'identifier certaines questions intéressantes au plan diachronique. Ainsi, j'ai relevé certaines similarités de structure entre parlers naish, sous des formes phonétiques qui, elles, diffèrent fortement : il existe une morphotonologie irrégulière partagée entre dialectes. Ces observations sont rapportées dans un article (Michaud 2011a) dont le résumé, reproduit ci-dessous, reste circonspect quant à l'origine diachronique exacte des similarités observés : « Bien qu'il ne soit pas encore possible de proposer de reconstruction... »



Numeral-plus-classifier phrases have relatively complex tone patterns in Naxi, Na (a.k.a. Mosuo) and Laze (a.k.a. Shuitian). These tone patterns have structural similarities across the three languages. Among the numerals from '1' to '10', three pairs emerge: '1' and '2' always have the same tonal behaviour; likewise, '4' and '5' share the same tone patterns, as do '6' and '8'. Even those tone patterns that are irregular in view of the synchronic phonology of the languages at issue are no exception to the structural identity within these three pairs of numerals. These pairs also behave identically in numerals from '11' to '99' and above, e.g. '16' and '18' share the same tone pattern. In view of the paucity of irregular morphology—and indeed of morphological alternations in general—in these languages, these structural properties appear interesting for phylogenetic research. The identical behaviour of these pairs of numerals originates in morpho-phonological properties that they shared at least as early as the stage preceding the separation of Naxi, Na and Laze, referred to as the proto-Naish stage. Although no reconstruction can be proposed as yet for these shared properties, it is argued that they provide a hint concerning the phylogenetic closeness of the three languages.

A l'initiative de Guillaume Jacques (spécialiste de diachronie), j'ai participé à deux travaux, l'un portant spécifiquement sur les langues naish, l'autre sur le transfert historique de nasalité entre un groupe de consonnes et la voyelle qui le suit. Ils sont présentés dans les paragraphes qui suivent (§1.4.1 et 1.4.2).

### 1.4.1. Etude de la phonologie historique des langues naish

Guillaume Jacques et moi avons réalisé un travail de reconstruction du sous-groupe naish du sino-tibétain, qui est paru dans la revue *Diachronica* (Jacques & Michaud 2011). Le résumé de ce travail est le suivant :

Le naxi, le na et le lazé sont trois langues dont la position précise au sein du sino-tibétain demeure un sujet de controverse. Nous défendons l'hypothèse selon laquelle elles partagent un ancêtre commun, le « proto-naish ». A la différence de langues conservatrices telles que le rgyalrong et le tibétain, qui possèdent des groupes de consonnes à l'initiale et des consonnes finales, naxi, na et lazé partagent une structure syllabique simple, conséquence d'une érosion phonologique poussée. L'interprétation des correspondances régulières entre ces langues requiert la formulation d'hypothèses au sujet de la structure phonologique du proto-naish. L'analyse, en partie guidée par des formes potentiellement apparentées dans les langues conservatrices, fait ressortir de nombreux cas de conditionnement phonétique de la voyelle par le lieu d'articulation de l'élément consonantique qui la précédait. Cette étude illustre le fait que des avancées importantes sont possibles dans l'étude de la phonologie historique de langues sino-tibétaines même très érodées.

Cet article propose une reconstruction de formes du stade proto-naish (dont les Annexes de l'article fournissent une liste) sur la base des correspondances entre formes attestées.

En marge de ce travail, nous avons exploité une source ancienne pour en tirer quelques indices au sujet de l'histoire contemporaine (diachronie à faible profondeur). Les vocabulaires de cinq langues d'Asie orientale publiés en 1903 par Charles-Eudes Bonin (Bonin 1903) sont transcrits de façon rudimentaire, selon les conventions orthographiques du français. Ils fournissent néanmoins des indices concernant la prononciation de ces langues peu avant 1900. Nous avons examiné deux des listes de Bonin (naxi et pumi) à la lumière de données plus récentes et plus systématiques, afin de déceler d'éventuelles indications sur des changements phonétiques. La liste de mots naxi fournit des indices concernant le degré de palatalisation des vélaires devant les voyelles fermées d'avant et la prononciation des voyelles /i/, /y/ et /o/. La liste de mots pumi révèle que le groupe /st-/ existait encore à l'époque dans le dialecte étudié. Ce modeste travail est paru en 2010 (Michaud & Jacques 2010).

### **1.4.2. Etude du transfert de la nasalité entre un groupe de consonnes initial et une voyelle**

La deuxième étude menée avec Guillaume Jacques est directement liée au travail de reconstruction centré sur les langues naish. Elle concerne l'étude diachronique des cas de transfert de nasalité entre un groupe de consonnes et une voyelle qui le suit. Le point de départ de cette étude était une observation née de la comparaison entre dialectes de la langue naxi : correspondance entre syllabes nasales type /hỹ/, /hĩ/, /hvỹ/, dans un dialecte, et /çy/, /çi/, /fv/ dans un autre (Michaud 2006c). Cette observation appelait une extension de l'étude à d'autres langues, car la correspondance ci-dessus ne permet en elle-même de remonter qu'à un stade où la syllabe possédait un élément nasal, sans qu'il soit possible de déterminer s'il était situé sur l'initiale (consonne ou groupe de consonne) ou sur la rime (voyelle). L'élargissement de ce travail a été mené en collaboration avec Guillaume Jacques, qui circule avec une aisance impressionnante dans l'abondante bibliographie de linguistique historique des domaines les plus divers. Des changements similaires ont été relevés dans plusieurs langues sino-tibétaines ainsi que d'autres familles. Des faits de langues amérindiennes ont paru particulièrement précieux pour affiner le modèle. Afin d'explorer cette piste de façon approfondie, des contacts ont été noués avec un collègue américain (contacté à la suggestion de Françoise Rose), qui a accepté de s'associer à nous, débouchant sur une publication conjointe (Michaud, Jacques & Rankin 2012) dans la revue

*Diachronica*. Les phénomènes décrits dans l'article, quoique bien attestés, sont relativement peu courants. En outre, la complexité des faits rend cet article relativement ardu, en dépit de nos efforts pour offrir au lecteur le meilleur confort de lecture possible. Néanmoins il s'agit d'un travail approfondi, dont il y a lieu d'espérer qu'il rende des services à ses lecteurs. Le résumé de l'article est le suivant :

Des données comparatives de plusieurs familles de langues montrent l'existence de transferts de nasalité entre un groupe de consonnes en position initiale de syllabe et la voyelle qui suit. Le passage en revue des exemples décrits à ce jour est complété par une nouvelle analyse de données sino-tibétaines. De prime abord, il semblerait que ce transfert puisse s'opérer dans les deux sens : de l'attaque consonantique vers la voyelle suivante – en tai-kadai, austroasiatique, sino-tibétain, niger-congo (kwa) et indo-européen (celtique) – et de la voyelle à la consonne précédente en sioux. L'examen des conditions d'apparition de ces changements révèle néanmoins une asymétrie. Le cas de figure le plus courant est que le transfert de nasalité s'opère de l'attaque consonantique vers la voyelle qui suit ; les cas que nous avons pu trouver d'un changement régulier dans la direction opposée proviennent tous de langues dans lesquelles les sons nasals connaissent l'une des restrictions suivantes : soit les consonnes nasales n'ont pas valeur de phonèmes (c.-à.-d. que leur apparition est déterminée par le contexte), soit l'opposition entre voyelles orales et nasales est neutralisée après les consonnes nasales (en faveur de voyelles nasales).

### 1.4.3. Autres travaux en phonétique historique

Outre les deux travaux réalisés avec Guillaume Jacques, j'ai rédigé un travail de synthèse au sujet de la monosyllabisation en Asie orientale. Les changements phonétiques présentent en effet des caractéristiques aréales ; de nombreuses langues d'Asie orientale ont subi une monosyllabisation et une forte érosion phonologique. Suite à un exposé (*invited talk*) à un colloque à Brême en 2009, le texte a été publié sous le titre *Monosyllabification: patterns of evolution in Asian languages* (Michaud 2012c). Cette synthèse reprend certains résultats de l'étude de cas portant sur les langues naish (Jacques & Michaud 2011), mais vise à atteindre un niveau de généralisation plus élevé. Elle est plus aboutie qu'une précédente synthèse, restreinte celle-là aux systèmes tonals : « Les systèmes de tons en Asie orientale : typologie, schémas évolutifs et modélisation » (Michaud 2011b).

Dans le domaine austroasiatique, Michel Ferlus a bien voulu m'associer, ainsi qu'une étudiante vietnamienne, à un travail au sujet d'un dialecte du centre-nord du Vietnam (Michaud, Ferlus & Nguyễn 2015). L'étudiante et moi l'avons aidé à mener l'enquête linguistique ; au moyen de quelques exemples, organisés d'après les catégories de la proto-langue (proto-vietique), Michel Ferlus a esquissé un tableau historique ; et nous avons rédigé et explicité ses observations, sous son contrôle. La poursuite de travaux en

linguistique historique des langues vietiques est l'un des projets à long terme de cette étudiante, dont je compte diriger la thèse en phonétique expérimentale à partir de l'année universitaire 2017-2018.

Enfin, les discussions au sujet de questions de phonétique historique ont une place centrale dans un travail éditorial : un projet de traduction de travaux d'André-Georges Haudricourt. Martine Mazaudon, Boyd Michailovsky et moi-même travaillons depuis 2009 avec une équipe internationale de traducteurs (spécialistes du domaine concerné) à la préparation d'un recueil des principaux travaux d'Haudricourt en traduction anglaise, sous le titre *Studies in the evolution of languages and techniques*. L'ouvrage est accepté pour publication par De Gruyter Mouton ; un contrat a été signé en 2013. Les éditeurs visent une finalisation vers 2018. Le texte de présentation est le suivant :

This book makes available, for the first time in English, a selection of the writings of André-Georges Haudricourt (1911-1996) on linguistics, ethnology, and the history of technology. Best known for his work on the evolution of languages, Haudricourt first trained in agronomy in Paris, and studied plant genetics in the laboratory of Nikolai Vavilov before beginning research in ethnobotany. A deep understanding of evolution and genetics, a functionalist perspective drawn from his interest in technology, and a firm belief that "science is one" inform his work.

His main articles, an unpublished item in linguistics, and a book excerpt on the history of the plough are translated here from Haudricourt's famously elliptical French by area specialists, with notes and epilogues to help readers appreciate their scientific and methodological relevance. The collection includes the seminal articles in which Haudricourt argued from comparative and functional linguistics and from cultural history that East Asian tone systems are not a common inheritance but arose in parallel, according to common principles, in languages originally without tone.

## 1.5. Méthodologie : de la collecte à l'archivage des données linguistiques

Documents joints au dossier : 2 articles :

Michailovsky, Boyd, Martine Mazaudon, Alexis Michaud, Séverine Guillaume, Alexandre François & Evangelia Adamou. 2014. Documenting and researching endangered languages: the Pangloss Collection. *Language Documentation and Conservation* 8. 119–135.

Niebuhr, Oliver & Alexis Michaud. 2015. Speech data acquisition: the underestimated challenge. *KALIPHO - Kieler Arbeiten zur Linguistik und Phonetik* 3. 1–42.

Venu à la phonétique après des études littéraires, je ne pouvais manquer d'être étonné par la différence de traitement de l'objet d'étude dans ces deux domaines. Dans les études littéraires, on apprend, par les exercices complémentaires que sont l'explication de textes et la traduction littéraire, à soupeser les conséquences des choix les plus menus : l'ajout ou le retrait d'une virgule, le remplacement d'un mot par un quasi-synonyme... Le respect de la lettre veut que chaque détail soit significatif. A l'inverse, il me semblait constater une certaine désinvolture dans le traitement des données phonétiques, lors de la collecte (protocoles expérimentaux trop peu spécifiques) ainsi que dans le traitement des données après récolte. Ma première communication à un colloque (Michaud 2002) était un plaidoyer « pour une charte de qualité des corpus ».

Les bases de données sonores abritées par les centres de recherches en phonétique sont paradoxalement assez peu fréquentées et peu structurées, si on les compare, par exemple, avec les bibliothèques universitaires. Les chercheurs et étudiants ont tendance à constituer leur propre corpus à mesure des besoins de leur recherche, plutôt que de raisonner en termes de patrimoine documentaire partagé. Les fonds d'archives sont peu connus, les grands corpus distribués sur Internet dépassent souvent les budgets du chercheur individuel, tandis que l'on peut enregistrer soi-même un corpus d'une qualité technique satisfaisante.

On voudrait souligner ici les limites de cette logique : il est illusoire de penser que l'on peut à tout moment créer le corpus dont on a besoin. Dans le cas des langues en danger, la mise en commun des données existantes est particulièrement nécessaire. Mais dans l'étude des grandes langues, le travail documentaire ne demande pas moins de sérieux.

Comme on l'imagine, ce cours de morale dispensé par un nouveau venu drapé dans des habits de bonne sœur à cornettes n'a guère recueilli l'adhésion des collègues. Une vérification au moyen des gadgets de bibliométrie des citations révèle que cette première publication n'a été citée que par un voisin de palier de l'époque au sein du laboratoire Lacito (Jacobson 2004). Mes convictions n'ont pas changé, et une collègue relevait en 2015 que le constat de 2002 concernant l'état des lieux de la documentation en

phonétique/phonologie aujourd'hui restait d'actualité ; en revanche, j'espère, sans en être vraiment sûr, que les travers de la communication programmatique de 2002 ne transparaissent plus que sous une forme atténuée dans les deux travaux récents joints au présent dossier, dont l'un concerne la collecte et l'analyse de données, l'autre leur archivage et diffusion.

## 1.5.1. Méthodes de collecte et analyse de données

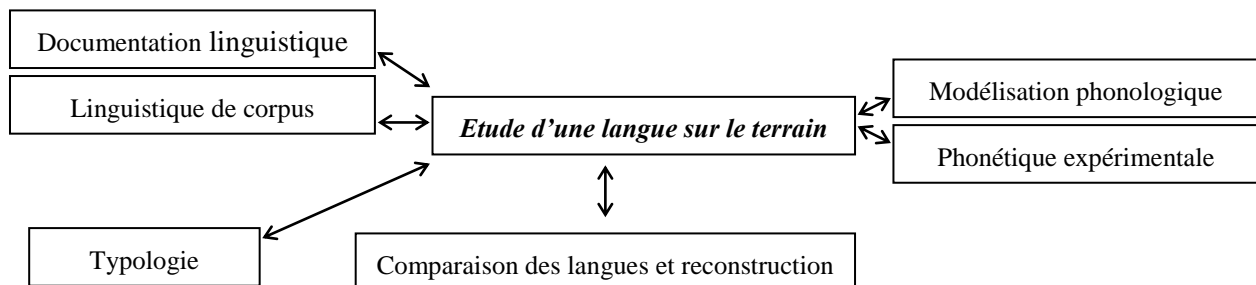
### 1.5.1.1. Principes

Depuis les débuts de mon travail sur le terrain, je m'efforce d'*amener le laboratoire sur le terrain*, dans les localités où sont parlées les langues concernées, plutôt que d'*amener les langues au laboratoire*. Les contraintes spécifiques de la linguistique de terrain peuvent, en pratique, se retourner en avantages.

Amener les instruments sur le terrain de l'enquête est de prime abord plus contraignant que de faire venir les informateurs au laboratoire de phonétique : sur le terrain, l'acoustique des lieux d'enregistrement n'est pas idéale ; visites de voisins et tâches domestiques interrompent fréquemment le travail ; les appareils de mesure sont fragiles et encombrants. En outre, il est fréquent que les mêmes consultants linguistiques participent aussi bien aux enregistrements qu'aux transcriptions et aux tests de perception, et cumulent le rôle de locuteur avec celui d'hôte et d'ami ; or la participation à l'enquête linguistique est susceptible de modifier certains aspects du comportement du locuteur, et la familiarité entre enquêteur et enquêté peut faire peser un doute sur l'objectivité des résultats.

Néanmoins, les précautions habituellement prises dans les laboratoires de phonétique (telles que : consignes écrites, rétribution fixée) n'atteignent pas nécessairement leur objectif : certaines données prises en laboratoire prêtent à critique, par exemple du fait de consignes prêtant à malentendus, ou du recours à des locuteurs habitant depuis longtemps à l'étranger (tandis que le travail de terrain permet d'étudier la langue là où elle *coule de source*, et offre également à l'enquêteur le loisir d'évaluer avec précision l'effet d'un éventuel bilinguisme sur les locuteurs qui travaillent avec lui). A l'inverse, les spécificités du travail de terrain peuvent se retourner en avantages. Les hypothèses suggérées par l'observation à l'oreille (lors de la transcription de récits, ou dans les échanges du quotidien) peuvent être soumises à vérification sans délai, par des expériences phonétiques ciblées.

Concevoir des expériences adaptées pour recueillir le sentiment linguistique de locuteurs peu familiers des tests de tous ordres s'avère non seulement réalisable, mais très utile pour faire avancer la réflexion. Au moyen d'un équipement de pointe autrefois réservé au laboratoire, il est également possible de réaliser, au fil des semaines, des corpus de grande taille (et d'excellente qualité). De longs enregistrements de parole continue (récits de vie, récits traditionnels, épisodes de la vie du village) ont été réalisés alors que le locuteur portait les électrodes d'un glottographe, ou un transducteur ferromagnétique (placé contre une narine pour l'étude de la nasalisation), cela sans difficulté, les locuteurs ayant eu le temps de s'accoutumer à ces techniques impressionnantes mais absolument sans danger. Les abondants matériaux recueillis peuvent être abordés à l'aide des outils (notamment statistiques) de la linguistique de corpus. La linguistique de terrain, *linguistique des langues*, offre une ouverture exceptionnelle sur diverses branches de la linguistique. Elle permet de relativiser les débats tels que celui qui oppose phonétique et phonologie, deux disciplines dont le tête-à-tête est peut-être trop exclusif pour être fructueux. Le tableau très schématique proposé ci-dessous vise à donner une idée de la diversité des approches qui nourrissent mon programme de recherche, et qui, en pratique, se complètent très utilement, sans bien sûr se confondre. (Les disciplines les plus importantes dans mon travail apparaissent en caractères gras.)



J'ai été sollicité de temps à autres pour présenter les méthodes employées dans mon travail de collecte et analyse de données : en particulier en 2012, pour un tutoriel commun avec Oliver Niebuhr à la conférence *Tonal Aspects of Language 2012* (à Nankin), pour lequel il a choisi le titre « Speech data acquisition: the underestimated challenge ». Notre espoir était que la combinaison de nos deux points de vue – lui plongé dans le quotidien des laboratoires de phonétique, moi ayant un pied dans la linguistique dite « de terrain » – donne une certaine largeur de champ à ce travail, publié en 2015 dans une anthologie de *working papers* de l'Université de Kiel (Niebuhr & Michaud 2015).

La question du public auquel on s'adresse, et de la « communauté » dans laquelle on se reconnaît, se pose de façon particulièrement aiguë lorsqu'on fréquente régulièrement des collègues spécialistes de diverses disciplines – en particulier : phonéticiens/phonologues, linguistes de terrain, et chercheurs en ingénierie linguistique. Les pratiques et les sensibilités dans le rapport aux données empiriques diffèrent d'une discipline à l'autre. Parmi les linguistes « de terrain », certains sont demandeurs d'informations concernant la prise de données phonétiques. En revanche, les phonéticiens ont rarement une vision patrimoniale des données qu'ils étudient, et peu d'entre eux se laissent convaincre qu'ils auraient fort à gagner, pour leur recherche, à investir plus de temps dans une réflexion approfondie au sujet des procédures d'élicitation, et dans une mise en forme et un archivage de leurs données dans les règles de l'art. Sans doute cela tient-il à la facilité apparente de la collecte des données en phonétique : l'illusion d'une quasi-fixité des langues que donne la continuité de langues nationales européennes telles que le français entretient l'illusion selon laquelle on pourrait à tout moment collecter les données dont on a besoin. Le linguiste qui étudie des langues sur le terrain, lui, est porté à prêter une haute valeur à des données collectées en surmontant des obstacles de toutes sortes : à ses yeux, « la collecte de données est une activité si prenante et si exigeante qu'elle invite le chercheur à valoriser ces données si durement acquises, et à les présenter dans des formats adaptés à un lectorat de spécialistes d'obédiences diverses » (Jacques 2014:2).

Même s'il ne paraît pas réaliste d'espérer un fort effet d'entraînement auprès des phonéticiens, du moins le travail réalisé avec Oliver Niebuhr offre-t-il une référence à laquelle renvoyer les étudiants, et une base de discussion au sujet des méthodes en phonétique expérimentale pour les collègues que cela intéresse.

### **1.5.1.2. Un outil logiciel : script pour l'analyse du signal électroglottographique**

Au cours de mes débuts dans l'analyse du signal électroglottographique (2001-2002), j'ai mesuré la nécessité de disposer d'un outil logiciel qui permette cette analyse pour des portions de signal courtes, celles que l'on rencontre ordinairement en parole, où le voisement est fréquemment interrompu pour la réalisation de séquences de phonèmes non voisés. Le script écrit par Nathalie Henrich pour la voix chantée (DECOM, script Matlab avec interface graphique) n'était pas approprié pour la voix parlée. Un collègue du laboratoire LIMSI, Tuan Vu-Ngoc, a bien voulu me communiquer les programmes en C qu'il avait écrits. La réflexion au sujet de l'outil a progressé en même temps que le



traitement de données, et suggéré des perspectives d'amélioration dont chacune représentait pour Tuan Vu-Ngoc un investissement de temps, à concilier avec ses tâches d'enseignement et ses obligations diverses, ce qui me causait une gêne croissante. Sur le conseil de Jacqueline Vaissière, je me suis mis à la programmation, sur la base de quelques rudiments acquis au lycée en option Informatique. Cela m'a permis de mieux comprendre le fonctionnement des outils logiciels, et de réaliser moi-même les modifications, expérimentations et améliorations qui me paraissaient souhaitables.

Le script Matlab que j'ai écrit pour l'analyse du signal électroglottographique a été baptisé PEAKDET (pour *peak detection*). Dans un premier temps, il a été hébergé sur des pages internet réalisées avec par Nathalie Henrich et Cédric Gendrot, et qui sont toujours en ligne à la date de rédaction du présent mémoire (<http://voiceresearch.free.fr/egg/index.html>).

Le script PEAKDET a été utilisé par plusieurs collègues, dont Daniel Recasens (Barcelone) et Marc Brunelle (Ottawa), avec de menus ajustements réalisés par des informaticiens recrutés dans le cadre de projets de recherche. Leurs versions modifiées de PEAKDET ont elles aussi été mises à disposition via le site voiceresearch.free.fr, avec leur accord. Cette mutualisation des outils a permis une économie de temps pour la petite communauté des utilisateurs de l'électroglottographe : les nouveaux utilisateurs ont pu se contenter d'ajustements mineurs, au lieu de nouveaux développements logiciels.

Lors du colloque Interspeech 2013, j'ai appris l'existence d'un repository Github qui accueille le code source d'algorithmes de traitement du signal (signal audio, et aussi des signaux présentant des similarités avec l'audio) : COVAREP (Degottex et al. 2014). Le responsable de ce repository a accepté d'héberger PEAKDET, ainsi que le script DECOM de Nathalie Henrich-Bernardoni. Après comparaison des versions de PEAKDET modifiées par les collaborateurs de Daniel Recasens et Marc Brunelle, le responsable a conclu que les différences ne concernaient que l'interface, laquelle n'est pas hébergée dans COVAREP, et n'a donc retenu qu'une seule des trois versions : le script PEAKDET « d'origine ».

La mise à libre disposition de cet outil reflète le même choix qui préside à la diffusion en libre accès des données de la Collection Pangloss, laquelle fait l'objet de la section qui suit.

### 1.5.1.3. Un questionnaire linguistique électronique

A l'occasion de la numérisation et mise en forme des données recueillies en Asie du Sud-Est par Michel Ferlus (voir §1.5.2.2), le questionnaire lexical de Michel Ferlus a été transformé en un document électronique se prêtant à certaines manipulations informatiques.

Cette liste numérotée vise à permettre aux chercheurs de naviguer entre les langues et les dialectes recueillis au fil des ans et sur tous les terrains d'Asie du Sud-Est. La première version a été élaborée par l'Ecole Française d'Extrême-Orient (EFEO), alors dirigée par Georges Cœdès, pour une vaste enquête lancée en 1938 et interrompue par la guerre en 1940. L'EFEO en a imprimé une quantité sous la forme de petits fascicules, distribués aux fonctionnaires envoyés en mission par l'administration coloniale (*Questionnaire linguistique*, Hanoi: Imprimerie d'Extrême-Orient, 1938). Ces fascicules, remplis avec plus ou moins de talent, constituent l'une des principales bases des travaux comparatistes réalisés par Haudricourt lors de son séjour à l'EFEO de Hanoi, à la fin des années 1940 (voir Haudricourt 1956:308).

Une version enrichie a été élaborée au laboratoire CeDRASEMI du CNRS (Centre de documentation et de recherche sur l'Asie du Sud-Est et le monde insulindien). Lucien Bernot a été la cheville ouvrière de cette amélioration qui a dû se faire entre 1960 et 1970. Cette version était décrite comme un "Questionnaire linguistique préparé conjointement par le Centre de documentation et de recherche sur l'Asie du Sud-Est et le monde insulindien (EPHE-CNRS, Paris) et le Département d'Asie du Sud-Est et d'Océanie de la School of Oriental and African Studies (University of London) en vue de l'établissement d'un Atlas ethnolinguistique de l'Asie du Sud-Est".

Michel Ferlus a re-tapé cette liste pour en faire des cahiers d'enquête commodes à remplir sur le terrain. Comme cette liste restait insuffisante pour une bonne utilisation linguistique, Michel Ferlus, au cours de ses enquêtes au Vietnam dans les années 90, a augmentée cette liste des nombres en *italique*, et a fait établir une traduction en vietnamien par Nguyễn Phú Phong ; cette liste a ensuite été remaniée par Trần Trí Dõi, qui a remplacé les mots de vocabulaire du dialecte du sud du Vietnam par le vocabulaire du nord du pays, plus adapté aux terrains que Michel Ferlus explorait avec lui à cette époque. Cette liste a circulé parmi les collègues et collaborateurs de Michel Ferlus. Elle a été enrichie de gloses en khmer par Frédéric Pain, en partie fondées sur la version de la liste CeDRASEMI-SOAS annotée en khmer par Marie Martin.

Le dépôt mis à disposition en ligne à l'adresse suivante : <https://halshs.archives-ouvertes.fr/halshs-01068533/> comporte la liste de vocabulaire au format tableur et un document de présentation trilingue, anglais-français-vietnamien. La préparation de la première version électronique (2014) a comporté les tâches suivantes:

- le document a été converti au format tableur
- une nouvelle numérotation a été établie, pour faciliter l'emploi de la liste par un identifiant unique assigné de façon continue de 1 à n
- des gloses en chinois ont été ajoutées ; les gloses en anglais ont été complétées ; et les gloses en vietnamien ont été intégralement révisées.

Lorsqu'un utilisateur passe du simple format tableur (liste de mot) à un format de dictionnaire en bonne et due forme, il peut conserver l'identifiant unique du mot, pour faciliter la comparaison avec d'autres langues pour lesquelles une enquête a été menée sur la base de la même liste. Michel Ferlus n'a pas souhaité apparaître comme premier auteur de cet outil informatique, notant le changement d'orientation entre la liste en français qui constituait son outil personnel et la liste multilingue créée par une équipe que j'animais. La version 1 a été déposée dans l'archive HAL-SHS sous les noms des principaux contributeurs (Pain et al. 2014) ; cela constitue un moyen de reconnaître leur contribution, même si la liste d'auteurs par elle-même n'indique pas précisément la contribution de chacun.

Des collègues en France et à l'étranger ont signalé que cet outil rendait service. En 2016, un jeune collègue, Ryan Gehrman, a spontanément proposé de coordonner un travail d'équipe qui a permis de revoir intégralement les gloses anglaises, ce qui a bien sûr été accepté avec gratitude. Minh-Châu Nguyễn, étudiante de Master, a également apporté des améliorations aux gloses vietnamiennes. Une deuxième version a alors été déposée (le système HAL-SHS permet la gestion de versions successives d'un même document). A l'heure actuelle (2017), Ryan Gehrman supervise une équipe qui va ajouter des gloses en birman, thai central, thai du nord et laotien. Il s'agit d'un travail effectué par des utilisateurs de la liste qui sont disposés à contribuer bénévolement à leur enrichissement. L'enrichissement par ajout de gloses dans d'autres langues et dialectes a vocation à se poursuivre.

## 1.5.2. Archivage et diffusion des données

### 1.5.2.1. Principes

L'archivage et la diffusion des données linguistiques ne sont guère considérés à l'heure actuelle comme faisant partie du cœur de métier des chercheurs et enseignants-chercheurs. L'enjeu est pourtant central : à mes yeux, ce qui est en jeu, c'est la réalisation de progrès cumulatifs dans le domaine de la documentation linguistique, pour le bénéfice de la recherche linguistique et des technologies vocales.

Cette évidence est rappelée régulièrement<sup>11</sup>, mais ce domaine souffre de plusieurs difficultés. Les efforts investis dans les tâches de collecte et mise en forme de données sont beaucoup moins valorisés institutionnellement que la publication de résultats scientifiques ; en partie pour cette raison, la collecte des données ne reçoit pas toujours l'attention qu'elle mérite, ce qui limite d'emblée la fiabilité des données et décourage leur réutilisation. En outre, la dialectalisation des formats hypothèque d'emblée l'interopérabilité. Dans ce contexte, il peut être nécessaire de rappeler une évidence parfois voilée par le primat accordé aux facteurs bibliométriques : la base empirique du travail est d'une importance considérable pour toutes les étapes ultérieures. Il paraît extrêmement prometteur de tenter un nivellement par le haut : poser de hautes exigences concernant la qualité des données expérimentales recueillies, de façon à ce que les données se prêtent au plus grand éventail possible d'exploitations et de questionnements. Les annotations de ces « données durables » se trouvent progressivement enrichies, de façon collaborative, au fil des utilisations et applications.

Un exposé sur ce thème a été présenté à la conférence IALP 2012 à Hanoi en novembre, en collaboration avec Andrew Hardie, de l'Univ. de Lancaster ; Séverine Guillaume, du laboratoire Lacito ; et Martine Toda, du laboratoire LPP (Michaud et al. 2012). Dans l'adoption de formats d'annotation, l'expérience acquise dans le cadre de la Collection Pangloss (anciennement « Programme Archivage du Lacito ») montre par exemple tous les bénéfices que l'on peut retirer de l'emploi de formats-pivots tels que XML, facilitant l'interopérabilité.

---

<sup>11</sup> Voir par exemple un billet de Mark Liberman, « (Not) trusting data » : <http://languagelog.ldc.upenn.edu/nll/?p=5590>

Je participe depuis 2001 aux activités de la Collection Pangloss du Lacito (anciennement Programme Archivage) : mise en forme, conservation et diffusion de documents linguistiques multimédia de langues rares. Il ne s'agit pas d'un rôle officiel, le responsable officiel de la Collection étant le directeur du laboratoire. Je consacre du temps et des efforts à des tâches variées, dont l'identification et le traitement documentaire de fonds de chercheurs (documents sonores de la donation René Gsell en 2001, et collections de Michel Ferlus de 2004 à 2016 : voir « Principale réalisation » en section suivante), l'accueil et l'encadrement de collaborateurs occasionnels (stagiaires, informaticiens en contrat à durée déterminée...), et l'aide ponctuelle à des collègues intéressés à verser des documents à l'archive. Selon l'état du document préparé par le chercheur, les tâches varient :

- conversion du format *texte simple* au format XML (structuré logiquement), par exemple pour le premier texte en langue tamang mis en ligne, avec gloses mot à mot : « La fille du moineau et les chevaux du roi ». Selon les cas, les outils utilisés sont XSLT ou Perl, avec plus ou moins de travail pour le paramétrage de l'outil et les retouches manuelles.
- numérisation de bandes magnétiques, par exemple la numérisation de bandes magnétiques UHER à la demande de Mme Denis Bernot : enquêtes au sujet de dialectes birmans. J'ai par la suite (2014) transmis ces données à Alice Vittrant, qui entamait un travail de numérisation et mise en ligne du fonds documentaire de Mme Bernot.
- ajout d'un alignement texte-son à des documents XML déjà prêts, par exemple pour les trois premiers textes en langue prinmi (pumi)

La communication par le biais de l'interface internet qui offre accès aux données est centrale ; je participe au fil des ans à la mise à jour des pages du site, et à la réflexion au sujet de l'identité et du positionnement du projet dans son contexte mondial. En 2012, l'archive du Lacito a adopté le nouveau nom « Collection Pangloss ». Le nom (trouvaille de Boyd Michailovsky) signifie étymologiquement « toutes les langues » ; on peut également y voir une allusion au gloses interlinéaires qui sont l'un des points forts de la collection. Le public français relèvera également l'allusion à *Candide*, et à la leçon de maître Pangloss : savoir cultiver son propre jardin – ce qui résume bien la philosophie de l'archive : mettre en ligne, en bon ordre, les données que nous collectons sur le terrain, et les exploiter, sans se laisser gagner par la fébrilité à la pensée des nombreuses langues qui sont en train de disparaître sans bruit avant d'avoir été documentées.

Un article collectif à douze mains a été réalisé pour présenter la Collection Pangloss (Michailovsky et al. 2014). Des communications ont également été présentées à des colloques, dont un exposé invité à Oriental-COCOSDA (Michailovsky, Michaud & Guillaume 2011) et un exposé à destination des phonéticiens aux Journées d'Etude de la Parole (Michaud et al. 2016). L'interface de consultation sur le site internet du Lacito est en cours de remaniement, facilitant notamment la consultation sous Android.

A l'Institut MICA à Hanoi, de novembre 2012 à juin 2016, j'ai eu à cœur d'établir entre MICA et le Lacito une collaboration profitable aux deux partenaires, dont les points forts sont complémentaires. Le Lacito est créateur de données d'une grande finesse, hébergées et distribuées dans une base de données de taille modeste mais d'architecture bien conçue. Ces données sont actuellement exploitées très en-deçà de leur potentiel, et ne sont pas systématiquement enrichies après la création de l'annotation linguistique traditionnelle (transcription, traduction et gloses). Pour sa part, le département « Speech Communication » de MICA est à la fois producteur de corpus qui peuvent atteindre des volumes importants, et utilisateur, ses membres étant au fait du potentiel qui s'ouvre lorsque l'on dispose d'une base de données dynamique. En combinant les exigences de précision (qualité linguistique) et de durabilité (réutilisabilité, interopérabilité...), cette collaboration entre MICA et le Lacito pourrait contribuer à l'excellence des ensembles documentaires constitués à MICA, et à accroître leur visibilité internationale. Ainsi, la préparation d'une communication de colloque (Nguyen et al. 2013) a fourni l'occasion de mettre en forme les données enregistrées et de les mettre en ligne dans les règles de l'art. Ces documents permettent de répliquer l'expérience que nous avons réalisée ; ils se prêtent également à d'autres finalités de recherche. L'ensemble documentaire mis en ligne à l'occasion de la préparation de cette communication de colloque est anecdotique, illustrant le tout-venant des données collectées ; mais sa préparation marque un transfert de technologie utile à l'Institut MICA. Dang-Khoa Mac, l'un des collègues de l'Institut, et co-auteur de cette communication de 2013, suivie en 2015 d'une étude dont il a pris l'initiative (Mac et al. 2015), a décidé de réaliser la mise en ligne d'un corpus audiovisuel qu'il avait enregistré lors de la préparation de sa thèse. Ces données sont disponibles depuis 2015 à l'adresse suivante :

<http://cocoon.huma-num.fr/exist/crdo/ark:/87895/1.17-498728>

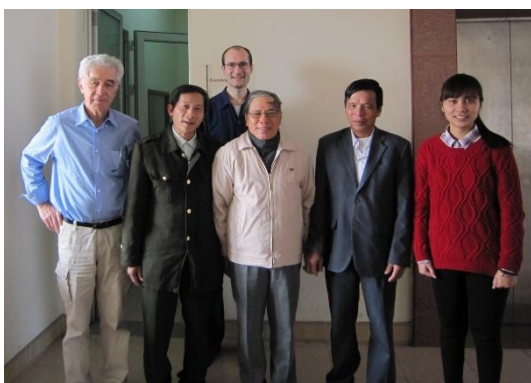
En 2013, Dang-Khoa Mac et moi avons été désignés comme co-responsables du projet AuCo (Audio Corpora) de l'Institut MICA. Le but de la collection AuCo est de recueillir les documents recueillis par les chercheurs au fil de leur activité de recherche. La

préparation, la réalisation et l'annotation d'enregistrements sont des tâches gourmandes en temps; au prix d'un léger surcroît de travail, il est possible de préparer les données de façon à ce qu'elles puissent être réutilisées par d'autres, pour des finalités variées (notamment: analyse phonétique/phonologique, traitement automatique de la parole, mais aussi didactique des langues/projets de revitalisation de langues en danger). La collection AuCo a vocation à contribuer à la documentation du patrimoine humain que représentent les langues du monde; elle a aussi vocation à faciliter les travaux de recherche interdisciplinaires associant ingénieurs et linguistes, autour de techniques communes.

En concertation avec Dang-Khoa Mac et d'autres collègues, plusieurs étapes ont été franchies. Une page internet trilingue exposant les objectifs du projet a été mise en ligne sur le site de l'Institut MICA. Un modèle de transfert de *copyright* (autorisation de diffusion selon une licence CreativeCommons) a été établi et mis en ligne sur cette page.

### 1.5.2.2. Principale réalisation : numérisation des données du linguiste Michel Ferlus

De septembre 2014 à février 2016, l'Institut MICA a réalisé un projet de numérisation intitulé « **DO-RE-MI-FA : Données des Recherches linguistiques de Michel Ferlus en Asie du sud-est** ». Ce projet concerne l'ensemble des documents linguistiques collectés par M. Michel Ferlus au cours de son activité comme « linguiste de terrain », de 1963 à 2003. Michel Ferlus est un spécialiste de la phonétique historique des langues d'Asie du Sud-Est.



*Michel Ferlus (à gauche) à l'Institut MICA en 2014, avec le Pr. Tran Tri Doi (Université Nationale), deux consultants linguistiques de la province du Quang Binh, et Minh-Châu Nguyễn. Au fond : A. Michaud. Photo Đặng-Khoa Mạc.*

Ces données inédites et irremplaçables proviennent de plus de 40 parlers jusque-là non documentés. Le fonds consiste en 200 heures d'enregistrements audio accompagnés de manuscrits et « tapuscrits ». Le travail a été réalisé à l'Institut de recherche international MICA (HUST-CNRS-Grenoble INP), situé au Vietnam, au cœur du domaine linguistique concerné. Archivage pérenne et diffusion en ligne –sous licence CreativeCommons– ont été réalisés dans le cadre de la Collection AuCo et la Collection Pangloss, sous l'égide de la Très Grande Infrastructure de Recherche Huma-Num. Le projet était financé par une subvention de 46.000€ de la Bibliothèque Scientifique Numérique du Ministère de l'Enseignement Supérieur et de la Recherche (<http://www.bibliothequescientifiquenumerique.fr/> BSN 5, Numérisation des données de l'enseignement supérieur et de la recherche, appel 2013).

L'objectif était de transformer le fonds de chercheur de Michel Ferlus en un ensemble documentaire dans les règles de l'art, pleinement exploitable, mis à la libre disposition de la communauté des chercheurs ainsi que d'un public plus étendu. Les tâches n'étaient pas de pure numérisation et catalogage, mais comportaient également un recollement entre ressources, une mise en forme des documents numériques, et des innovations dans leur affichage en ligne. S'agissant de données « de chercheur », il fallait réaliser une mise en valeur qui puisse convaincre un public de chercheurs. Le projet s'inscrivait dans la perspective de permettre l'émergence de nouvelles pratiques de recherche en ouvrant ce fonds à diverses utilisations, dont une exploitation par les outils d'analyse du signal et d'analyse de corpus linguistiques.

Le projet avait deux partenaires institutionnels : l'Institut MICA à Hanoi, porteur du projet ; et le Lacito en région parisienne, point d'entrée vers l'archivage pérenne, et interface de consultation. Le projet DO-RE-MI-FA regroupait dix personnes (sans compter sept stagiaires, présents pour des durées de 2 à 3 mois). Parmi elles, seulement quatre étaient des « permanents » de l'une ou l'autre des deux institutions partenaires (Séverine Guillaume au Lacito, et Dang-Khoa Mac, Ngoc-Diep Do et Alexis Michaud à MICA). Deux personnes travaillaient à distance, l'une de Toulouse (Matthew Deo), l'autre de Doha (Thu-Ha Pham) ; parmi les collaborateurs dont les tâches étaient réalisées dans les locaux de l'Institut MICA, un (Frédéric Pain) est ensuite retourné en Belgique, une (Hang Dinh) a un emploi à l'Institut de Linguistique de Hanoi, et une (Minh-Chau Nguyen) était étudiante en Master à l'Université Nationale de Hanoi.



Les outils logiciels ont été préparés par les chercheurs et ingénieurs en informatique du projet (Mac Dang-Khoa, Do Ngoc-Diep et Matthew Deo) en concertation avec Séverine Guillaume et les divers participants concernés. Les métadonnées et annotations ont été préparées à l'Institut MICA, vérifiées et transmises par le coordinateur à Séverine Guillaume pour vérification technique et dépôt. L'administration (préparation des contrats, suivi des dépenses...) était assurée par le coordinateur et Ngoc-Diep Do, qui suivaient le projet avec la Secrétaire de direction de l'Institut.

Le projet se caractérisait par la grande diversité de l'équipe, des compétences, des tâches, et des localisations géographiques. Ce partenariat « en archipel » a conduit l'équipe à adopter des méthodes que Matthew Deo nous a appris à nommer « méthodes agiles » : méthodes de développement logiciel, mais aussi de gestion. Le principe est le suivant. L'objectif de départ est fixé : en l'occurrence, il figure dans le projet déposé auprès de la Bibliothèque Scientifique Numérique. Les participants se réfèrent au coordinateur lorsqu'il faut convenir d'un ajustement au projet, ou à la répartition des tâches ; pour le reste, chacun a la responsabilité d'un petit domaine où il possède plus de connaissances que les autres, et a la charge de faire progresser sa partie du projet. A titre d'exemple, le développement et le déploiement de l'outil EASTLing (pour la création et l'affichage en ligne de documents multimédia comprenant un manuscrit affiché en mode image, et synchronisé avec un fichier audio contenant le texte lu) était une tâche complexe qui concernait un outil d'édition et un outil de lecture, et qui nécessitait leur bonne intégration à l'architecture existante de la Collection Pangloss. Ce travail a essentiellement eu lieu en dialogue entre Matthew Deo (auteur-concepteur) et Séverine Guillaume (responsable de la Collection Pangloss). Le coordinateur recevait en copie les échanges de courriels et pouvait, à l'occasion, apporter un complément d'information ou des suggestions.

Séverine Guillaume, ingénieur responsable de la Collection Pangloss, a fait bénéficier le projet de ses qualités exceptionnelles de souplesse et de disponibilité : elle a adapté son calendrier à celui des réalisations des divers participants. Loin de regretter les tâches additionnelles que lui amenait le projet, elle a constamment encouragé les uns et les autres, et salué chacun des dépôts comme un progrès de l'entreprise de documentation linguistique qu'elle orchestre.

Au final, l'intégralité des données audio concernées par le projet ont été numérisées. La numérisation des cassettes est en principe une tâche de routine. Mais s'agissant de cassettes compactes anciennes, ce travail demande une certaine expérience et

un certain doigté. Familier des tâches de numérisation par mon expérience du travail au sein de la Collection Pangloss depuis 2001, j'ai réalisé moi-même la numérisation dans son intégralité.

La mise en ligne en libre accès intégral, sans restriction, a été achevée dans les 18 mois suivant le début de mise en œuvre, conformément aux demandes de la Bibliothèque Scientifique Numérique. N'ont été retranchées que deux langues vietiques, pour lesquelles Michel Ferlus a souhaité différer l'ouverture au public, du fait qu'il travaille actuellement à l'analyse de ces données.

L'enrichissement de ces données par une annotation multilingue s'appuyant sur les notes de Michel Ferlus constitue une entreprise pour le moyen/long terme, indissociable de la formation de la jeune génération des chercheurs dans ce domaine scientifique. Cet immense chantier a été bien engagé dans le cadre du projet, avec la réalisation d'annotations pour la plupart des documents arem, mường, củi chām, ainsi que pour une trentaine de documents de langues tai-kadai, par des personnes dont certaines contribueront à prendre la relève du travail de recherche de Michel Ferlus. Dans le cadre du projet, 6 stagiaires de l'Université Nationale du Vietnam ont été accueillis pour 2 mois à l'Institut MICA. Ce stage compte comme stage de fin d'études universitaire (4e année : au Vietnam, le diplôme universitaire équivalent de la Licence du système européen « L-M-D » s'obtient en 4 ans). Parmi eux, trois continuent en Master, avec le projet de continuer jusqu'à la thèse. L'Institut MICA joue un rôle d'« accélérateur », et d'interface avec le monde international de la recherche, pour le plus grand profit de ces étudiants de linguistique. Sur le versant français, nous avons reçu une candidature spontanée à un stage bénévole dans le cadre du projet DO-RE-MI-FA. Julien Heurdier, Docteur en Sciences du langage de l'Université Paris 3, a pris contact en mars 2015 pour proposer de contribuer à mi-temps au projet DO-RE-MI-FA, à titre bénévole, de novembre 2015 à janvier 2016. L'objectif pour lui était de diversifier sa connaissance du monde de la recherche, et d'apporter une contribution à un projet qu'il jugeait utile et motivant. Il a réalisé (en collaboration étroite avec le coordinateur) le catalogage et les pages internet pour le khmer, le khamou, et diverses langues austroasiatiques.



*Le Pr. Tri-Doi Tran et A. Michaud à l'Institut MICA, 2015*



**L'ESPACE**  
**INSTITUT FRANÇAIS**  
HANOI

**LA COLLECTION AU CO : UNE ARCHIVE OUVERTE POUR LA CONSERVATION ET L'ÉTUDE DES LANGUES DU VIETNAM ET DES PAYS VOISINS**

**ÂU CO : CƠ SỞ DỮ LIỆU ÂM THANH ĐỂ BẢO TỒN VÀ NGHIÊN CỨU CÁC NGÔN NGỮ Ở VIỆT NAM VÀ CÁC NƯỚC LÃNG GIỀNG**

CONFÉRENCE HỘI THẢO



Enquête linguistique au Nghệ An, 2014. (Photo : AM)  
Diễn đàn ở Tương Dương, Nghệ An, 2014. (Ảnh: AM)

A l'heure d'une mondialisation qui conduit à une extinction massive des langues, une des tâches des "humanités numériques" consiste à utiliser les nouvelles technologies pour la conservation pérenne des langues menacées. La Collection AuCo regroupe et met en ligne des documents linguistiques de langues peu décrites, pour la plupart sans écriture. L'exposé offrira un parcours à travers l'extraordinaire richesse linguistique du Vietnam; il présentera les enjeux du projet, pour la recherche scientifique, mais aussi pour la conservation du patrimoine culturel.

Avec la participation du Pr. Tran Tri Doi, du Département de linguistique de l'Université Nationale de Hanoi, et d'Alexis Michaud, chercheur à l'Institut international de recherche MICA, HUST-CNRS.

Toàn cầu hóa dẫn đến sự tuyệt chủng của hàng loạt các ngôn ngữ trên thế giới. Một trong những nhiệm vụ của "Khoa học Nhân văn Điện tử" (Digital Humanities) là khai thác tiềm năng của công nghệ thông tin để bảo tồn bền vững các ngôn ngữ đang bị đe dọa. Cơ sở dữ liệu Au Co thu thập và lưu trữ các bản thu âm các ngôn ngữ trên lãnh thổ Việt Nam và cả các nước láng giềng, bao gồm những ngôn ngữ đang có nguy cơ tuyệt chủng và những ngôn ngữ nghèo dữ liệu. Buổi hội thảo này sẽ là một cuộc hành trình khám phá sự phong phú của các ngôn ngữ ở Việt Nam, và cũng sẽ phân tích các thách thức của dự án trong nghiên cứu khoa học và trong việc bảo tồn các di sản văn hóa.

Với sự tham gia của các diễn giả: GS.TS.Trần Trí Đới, Khoa Ngôn ngữ học, Trường Đại học Quốc gia Hà Nội; và Alexis MICHAUD, nghiên cứu viên tại Viện nghiên cứu quốc tế MICA, ĐH Bách Khoa Hà Nội - CNRS.

**09h00 20.09** (Samedi | Thứ 7)

Auditorium de l'Espace | Hội trường Trung tâm Văn hóa Pháp  
Traduction consécutive | Dịch tiếp nối Pháp - Việt  
Entrée libre | Vào cửa tự do





INSTITUT FRANÇAIS DE HANOI - L'ESPACE Trung tâm Văn hóa Pháp tại Hà Nội  
24 Tràng Tiền, Hà Nội. Tél: 39 36 21 64. Fax: 39 36 21 65. contact@espace-ccfhanoi.org. www.ifhanoi-lespace.com

*Annonce d'un exposé à l'Institut français de Hanoi (centre culturel français) en 2014 au sujet des archives linguistiques*

Etant donné les moyens limités de l'équipe du projet, il a paru souhaitable de ne pas introduire de redondances dans les tâches de développement d'interfaces spécialisées pour la consultation des documents. C'est donc la Collection Pangloss qui constitue l'interface de consultation pour les données du projet DO-RE-MI-FA. On a fait l'économie d'une réplique de cette interface sur le site internet de l'Institut MICA : celle-ci demanderait une maintenance continue (24/7) pour la continuité du service, laquelle est indispensable pour un bon référencement par OLAC et d'autres moteurs de recherche et centres d'indexation qui « moissonnent » régulièrement les métadonnées des Archives ouvertes. La solidité des bases mises en place par Pangloss/Cocoon, fruit de choix

opportuns dans les formats et les outils (Dublin Core, OLAC...), est reconnue internationalement : Cocoon est systématiquement dans les 5 premières archives du classement OLAC (<http://www.language-archives.org/metrics/compare>). Cela rejillit positivement sur les projets qui s'intègrent dans le cadre de cette Archive ouverte.

Les langues concernées par le projet apparaissent dans la liste des langues de la Collection Pangloss, dans la section « Asie ». Les documents peuvent être consultés en ligne, à l'aide d'un navigateur, sans qu'il soit besoin d'installer d'éléments logiciels supplémentaires : le HTML5 permet en effet la lecture de contenus multimédia sans avoir à faire appel à des logiciels propriétaires tels que QuickTime.

**Vocabulary list for Quang Binh dialect of the Arem language as spoken in the district of Bo Trach,**  
Language : Arem

Researcher(s) : Ferlus, Michel Speaker(s) : Đinh Óc; Đinh Châu; Dinh, Oc; Dinh, Chau

0:00 47:31  Continuous playing:

**Transcription**  
 Phonologic Notes

**Translation**  
 fr  en  cn  others

W1	<input type="checkbox"/> <input type="checkbox"/>	tləj	trời	sky	天	ciel	megh	me:c	mè:k
W2	<input type="checkbox"/> <input type="checkbox"/>	tləj	trời	sky	天	ciel	megh	me:c	mè:k
W3	<input type="checkbox"/> <input type="checkbox"/>	tləj	trời	sky	天	ciel	megh	me:c	mè:k
W4	<input type="checkbox"/> <input type="checkbox"/>	tləj	trời	sky	天	ciel	megh	me:c	mè:k
W5	<input type="checkbox"/> <input type="checkbox"/>	braw <sup>2</sup> p <sup>h</sup> æ:ŋ <sup>2</sup>	trời trong	clear sky	晴天	clair, ciel clair			
W6	<input type="checkbox"/> <input type="checkbox"/>	braw <sup>2</sup> p <sup>h</sup> æ:ŋ <sup>2</sup>	trời trong	clear sky	晴天	clair, ciel clair			
W7	<input type="checkbox"/> <input type="checkbox"/>	braw <sup>2</sup> jam	(trời) râm	dark sky	阴天	sombre, ciel sombre			
W8	<input type="checkbox"/> <input type="checkbox"/>	braw <sup>2</sup> jam	(trời) râm	dark sky	阴天	sombre, ciel sombre			
W9	<input type="checkbox"/> <input type="checkbox"/>	mèt karæ:ŋ	mặt trời	sun	太阳	soleil	thnai	l'ŋaj	l'ŋáj ʔej

*Un document du projet DO-RE-MI-FA, en ligne sur l'interface de la Collection Pangloss*

Toutes les nouvelles ressources déposées sont répertoriées par OLAC dans les jours qui suivent, et également par Ethnologue, qui fournit un lien dynamique vers le catalogue OLAC. Voir par exemple, pour la langue tai don (tai blanc) : la page « tai don » dans l'inventaire de langues *Ethnologue* (<https://www.ethnologue.com/language/twh>) offre un lien vers les ressources OLAC dans cette langue. Les ressources affichées dans la rubrique « Primary texts » ont été déposées dans le cadre du projet DO-RE-MI-FA.

Les ressources sont également indexées par les moteurs de recherche généralistes, sans qu'aucune action particulière n'ait été menée pour aider à leur référencement.

Pour ce qui est des transcriptions, le mode opérationnel actuel est le suivant : les manuscrits de Michel Ferlus sont remaniés un par un, et leur contenu publié, soit par l'auteur lui-même, soit par un étudiant-chercheur (doctorant) intéressé à reprendre le flambeau pour l'étude d'une langue en particulier, et ayant une certaine familiarité avec la langue (telle qu'elle est parlée aujourd'hui). La justification de ce choix est qu'un utilisateur qui est lui-même chercheur, engagé dans l'étude de la langue-cible, sait tirer le meilleur parti des notes de terrain, et redresser de lui-même les inévitables petites erreurs ou approximations dans la notation.

Ce cas de figure se présente actuellement pour la langue *mừng* : une étudiante vietnamienne, participante au projet, aborde, pour ses études doctorales, l'étude de la phonétique/phonologie de la langue *mừng*. La base empirique de son travail consiste d'abord en une enquête qu'elle réalise elle-même de point en point, sur le terrain. Au stade du doctorat et au-delà, elle pourra faire usage des notes de Michel Ferlus. A terme, forte d'une bonne compréhension de la langue, elle pourra toiletter les notes de Michel Ferlus, et compléter les documents actuellement en ligne (avec traduction alignée, mais sans transcription) par leur transcription en Alphabet Phonétique International.

Ce mode de fonctionnement peut paraître extrêmement contraignant. Il se peut qu'il faille attendre plusieurs années pour que se trouve un lecteur qui réponde à l'exigence d'une bonne connaissance préalable du domaine linguistique concerné. Pour autant, cette perspective n'est nullement utopique : les progrès sont bien réels, et des collègues intéressés au sein de la communauté universitaire se manifestent plus fréquemment qu'il n'était initialement prévu dans les objectifs modestes du projet.

Deux conclusions me paraissent ressortir de ce projet. Tout d'abord, la faisabilité d'un projet de documentation mené par des chercheurs. Il n'est pas impossible à un groupe de chercheurs, au sein d'une unité de recherche, de mener un projet documentaire, et que ce travail apporte un bénéfice pour la recherche. D'autre part, l'importance du lien entre les documents et les recherches qui ont motivé leur collecte. Les documents concernés par le projet DO-RE-MI-FA ont été collectés comme aide-mémoire par le linguiste dans le cadre de son travail, sans intention de conservation. Les numériser et les mettre en ligne « en masse » reviendrait à projeter dans l'espace numérique un fouillis d'enregistrements privés de leur contexte. Pour éviter de créer du désordre, il est particulièrement utile de recueillir auprès de l'auteur informations, documentation, et conseils pour la poursuite du travail de mise en forme et d'analyse. Ce dialogue peut jouer un rôle de stimulation auprès du

chercheur, l'amenant à reprendre des notes anciennes, pour en effectuer la saisie et le toilettage, ou pour décider de les confier à un tiers. A mes yeux, archivage et valorisation de corpus sont à intégrer à l'activité des unités de recherche, et non à « sous-traiter » à des spécialistes extérieurs.

## 2. Orientations futures

C'est une chance extraordinaire de pouvoir se consacrer à ses passions. Depuis mon entrée au CNRS en 2006, je goûte avec intensité le bonheur de bénéficier d'un statut qui me met « à l'abri du vent » pour mon existence entière, dans un contexte où nombre de collègues d'autres pays sont sous l'épée de Damoclès d'un non-renouvellement de leur contrat de travail. En pratique, ce statut laisse au chercheur la latitude de déterminer l'orientation de ses travaux.

Pour ne pas gaspiller cette chance extraordinaire, il est nécessaire de fixer des priorités : déterminer des objectifs à moyen et long terme, et m'efforcer de leur réserver la part de mon temps qui est nécessaire pour les atteindre. Après mon entrée au CNRS, l'incertitude quant à la façon dont serait meublées les journées au laboratoire a été levée bien vite : j'ai pu constater avec quelle rapidité le quotidien se remplit. Les sollicitations sont nombreuses, pour des activités d'une grande diversité. Les rubriques du formulaire de « Recueil d'Informations pour un oBServatoire des Activités de reCherche en SHS » (RIBAC), renseigné chaque année par les chercheurs en Sciences humaines, en donnent un aperçu : Activités de recherche et d'enseignement (collecte de données, publications scientifiques, enseignement et encadrement, congrès, colloques et tables rondes) ; responsabilités collectives ; expertises scientifiques ; activités éditoriales ; activités de transfert et de valorisation ; sites Web, Blogs, Wikis...

Au vu de l'expérience des dix années passées, mon choix pour les deux ou trois décennies qui me séparent de la retraite consiste à mettre l'accent sur les deux groupes de langues qui constituent désormais ma spécialité : les langues naish (famille sino-tibétaine), d'une part, et les langues vietiques (famille austroasiatique), d'autre part. Les orientations futures sont présentées ci-dessous en perspective aréale, puis de façon thématique.

## 2.1. Perspective aréale

### 2.1.1. Domaine des langues naish (sino-tibétain)

Dans l'étude des langues du groupe naish, une étude détaillée de chaque variété n'est pas à la portée d'un individu : il est indispensable de travailler en équipe. A l'heure actuelle (2017), le cercle des chercheurs qui peuvent se consacrer à ces langues est particulièrement restreint. Je souhaite contribuer à faire vivre ce domaine de recherche en dirigeant des travaux d'étudiants et en menant des projets communs avec des chercheurs confirmés. Ces travaux et projets peuvent consister, soit en un élargissement des recherches à de nouveaux dialectes, soit en travaux d'approfondissement au sujet de thèmes particulièrement intéressants concernant des dialectes parmi ceux qui ont déjà fait l'objet d'une description.

Pour prendre un premier exemple de travaux d'approfondissement, j'aimerais explorer avec des outils de phonétique expérimentale les phénomènes de nasalité dans divers dialectes naxi. Ce travail prendrait pour point de départ les conclusions d'une étude de linguistique comparée (diachronique) : les syllabes à initiale glottale et à voyelle nasale du dialecte naxi de Fengke proviennent d'un transfert de nasalité d'un ancien groupe de consonnes en position initiale de syllabe à la voyelle qui suit (Michaud, Jacques & Rankin 2012). Ce cadrage diachronique fournit un point de départ prometteur pour l'étude expérimentale synchronique, laquelle peut en retour contribuer à affiner la modélisation diachronique. Un travail dans ce sens a été entamé par une étudiante de Master, Yanchen Shuai (Université Grenoble Alpes). Pour commencer à se familiariser avec les faits du naxi de Fengke, Yanchen Shuai a analysé des enregistrements que j'avais réalisés en 2004 ; ceux-ci ont fourni la matière d'un mémoire de M1 dirigé par Nathalie Vallée (qui m'a fait l'honneur d'être co-directeur) intitulé « Étude acoustique exploratoire de la nasalité en naxi, langue sino-tibétaine » (soutenance : juin 2016). Yanchen Shuai s'est ensuite rendue sur le terrain au village de Fengke à l'été 2016. Actuellement (2016-2017), elle poursuit en M2 ces recherches, qui pourraient donner matière à une thèse de phonétique expérimentale au sujet de la nasalité. L'emploi d'outils de phonétique expérimentale « de pointe » permettrait d'obtenir des résultats d'une grande finesse ; le choix de les appliquer à la langue naxi, dont le système phonologique est éloigné de celui des langues couramment étudiées avec ces

techniques, est garant d'observations originales, qui seraient accueillies favorablement dans le cercle des phonéticiens-phonologues spécialistes de la nasalité.

Je participe également à l'encadrement des recherches doctorales de Likun He à l'Université centrale des Nationalités (Minzu University, Pékin). Likun He m'avait contacté en 2013 : locuteur natif de la langue naxi, il souhaitait s'engager dans des études doctorales et cherchait une orientation de recherche. Il a été accueilli à deux reprises à l'Institut MICA à Hanoi pour des stages, en 2014 et 2016, ce qui a permis de consolider sa formation de base, notamment à l'occasion d'un travail élémentaire d'inventaire phonologique de sa langue maternelle (Michaud & He 2015). Les échanges au fil des mois lui ont également permis de formuler un projet doctoral précis et réaliste. Il a choisi de prendre pour point de départ de ses recherches les hypothèses formulées dans deux articles que j'ai publiés au sujet des phénomènes de réduction syllabique en naxi et de leurs conséquences sur le système tonal (Michaud 2006a; Michaud & He Xueguang 2007). Son niveau d'anglais étant trop limité pour qu'il puisse envisager un doctorat dans cette langue, Likun He prépare son doctorat en Chine. A l'heure actuelle, ma participation à l'encadrement de son travail est informelle ; c'est sa directrice de thèse, Mme le Pr. Yan Liu (刘岩), qui décidera si elle juge opportun ou non de m'y associer officiellement, sous la forme d'une co-direction ou co-tutelle de thèse. En tout état de cause, ma collaboration avec Likun He va sans doute continuer et s'intensifier au fil des ans. Le choix qu'il a fait de se spécialiser en phonétique expérimentale paraît promettre de belles réalisations communes à venir, combinant un cadrage dialectal judicieux (choix de phénomènes qui présentent un intérêt particulier au plan phonétique-phonologique) avec des méthodes expérimentales *dans les règles de l'art*.

Un troisième projet doctoral concerne Yanjuan Mu, locutrice native de la langue naxi. Nous sommes en contact depuis 2011. En 2015, l'Université où elle enseigne (l'Université des Nationalités du Sud-Ouest, Chengdu, Chine) lui ayant proposé un aménagement de son service d'enseignement qui lui permette la préparation d'un doctorat, Yanjuan Mu a souhaité que je l'aide à formuler un projet de thèse. Son Université d'origine demandant à ce que le sujet soit inscrit dans un département d'Histoire, le choix s'est porté sur un sujet d'histoire des sciences et d'épistémologie. Je ne suis pas officiellement co-encadrant de sa thèse, mais nous sommes en contact suivi et des collaborations pourront se poursuivre après l'obtention par Yanjuan Mu de son Doctorat à l'Université des Nationalités du Sud-Ouest (Chengdu, Chine). De mars 2016 à février 2017, Yanjuan Mu a



été accueillie au laboratoire Lacito (avec un financement de son pays d'origine), où elle a pu travailler à ses recherches dans de bonnes conditions.

De façon informelle, je dispense des conseils à Emmanuelle Laurent, doctorante en anthropologie au Lacito, qui étudie la question de la parenté et de la transmission culturelle chez les Naxi. Bien que sa thèse ne concerne pas directement les Sciences du langage, son travail présente de nombreux points de contact avec le mien, à commencer par la collecte de documents fiables (qui, s'agissant d'une langue non écrite, sont des enregistrements audio et vidéo), leur transcription, et leur analyse approfondie. Pour le linguiste, les enregistrements qu'elle recueille dans le dialecte du village de Wumu constituent des documents d'une grande valeur, au sujet d'un dialecte qui à ma connaissance n'avait pas encore fait l'objet de telles collectes.

Je suis en outre en contact suivi avec Zihe Li (docteur de l'Université de Pékin en 2013), qui souhaite poursuivre et étendre ses enquêtes dialectologiques. Celles-ci constituent la base de ses recherches diachroniques ; il les conçoit aussi comme contribution à la documentation linguistique, objectif qui nous est commun. Selon l'évolution de sa carrière, il est envisageable que nous co-dirigions des travaux d'étudiants chinois, certains inscrits en Chine, d'autres en France.

En vue d'une bonne couverture dialectale du domaine des langues naish, plusieurs parlars appelleraient (à mes yeux) une monographie complète pour leur description et leur analyse : Fengke 奉科 et Labai 拉柏, sur les rives du fleuve Yangtsé ; Wujiao 屋脚 ; et Xiangjiao 项脚, dans le comté de Muli. Une collègue anthropologue, Pascale-Marie Milan, signale qu'en comparaison du village où j'ai travaillé, le déclin de la langue na est moins rapide dans les hameaux les plus reculés. Pour prendre l'exemple de Wujiao 屋脚, les difficultés matérielles de cette localité, telles que l'absence de réseau électrique et le climat plus rude que dans la plaine, n'attirent pas l'explorateur frileux que je suis. J'aime transporter sur le terrain certains éléments du laboratoire du phonéticien (micros, électroglottographe...) ; pour moi, les aléas tels que les coupures de courant – même de plusieurs jours – sont gérables, mais l'absence totale de réseau électrique m'aurait privé du confort informatique dont j'ai pris l'habitude dans mon travail. Au fil des ans, ces destinations excentrées deviennent moins inconfortables, et conseiller Wujiao ou Xiangjiao à des étudiant(e)s qui souhaiteraient pratiquer un terrain en immersion ne paraît ni cruel ni

pervers<sup>12</sup>. Pour des candidat(e)s moins intrépides, il existe des localités plus faciles d'accès, comme Fengke.

## 2.1.2. Domaine des langues vietiques (austro-asiatique)

Mon intérêt pour le domaine des langues vietiques tient essentiellement à la conjonction de deux facteurs. L'un tient à la grande richesse, au plan phonétique, des tons du vietnamien de Hanoi (langue que j'ai apprise de 1999 à 2001) et des parlers *muòng* qui sont ses plus proches parents. L'étude de ces parlers, dans leur diversité, a beaucoup à apporter à la typologie tonale. Le second facteur est le souhait de contribuer à maintenir une continuité dans le domaine des études vietiques, en prolongeant l'œuvre de Michel Ferlus (sans prétendre l'égaliser ni la dépasser). Michel Ferlus est un spécialiste de la phonétique historique des langues d'Asie du Sud-Est : non seulement les langues vietiques, et les langues austroasiatiques dans leur ensemble, mais aussi les langues tai-kadai. Michel Ferlus a apporté une contribution d'une importance bien connue des spécialistes ; en revanche, il n'a pas dirigé de travaux d'étudiants. Nos échanges à partir de l'an 2000 l'ont progressivement convaincu de l'opportunité de faciliter la relève dans son domaine par l'archivage et la diffusion des données qu'il a collectées de première main depuis 1963. La numérisation et la diffusion de ses données a été entreprise à partir de 2013, et a reçu un financement de la Bibliothèque Scientifique Numérique (Ministère de l'Enseignement supérieur et de la Recherche) : voir §1.5.2.2. Au fil de leur participation aux tâches de mise en forme, certains étudiants ont été confirmés dans leur vocation de devenir linguistes. L'expérience de ce projet m'a convaincu que je pouvais faire œuvre utile dans le domaine

---

<sup>12</sup> Qualificatif repris à la description par Claude Lévi-Strauss de son premier départ pour les Amériques, évoquant le « caprice un peu pervers » de l'enseignant qui envoyait ses étudiants sur un terrain qui leur était inconnu. Ils se trouvaient réunis pour un banquet de départ « dans un hôtel de l'avenue qui s'appelle aujourd'hui Franklin-Roosevelt ; demeure inhabitée, ou, pour l'occasion, un traiteur était venu deux heures auparavant installer son campement de réchauds et de vaisselle, sans qu'une aération hâtive ait réussi à purger l'endroit d'une odeur de désolation. Aussi peu habitués à la dignité d'un tel lieu qu'au poussiéreux ennui qu'il exhalait, assis autour d'une table trop petite pour un vaste salon dont on avait tout juste eu le temps de balayer la partie centrale effectivement occupée, nous prenions pour la première fois contact les uns avec les autres, jeunes professeurs qui venions à peine de débiter dans nos lycées de province et que le caprice un peu pervers de Georges Dumas allait brusquement faire passer de l'humide hivernage dans les hôtels meublés de sous-préfecture, imprégnés d'une odeur de grog, de cave et de sarments refroidis aux mers tropicales et aux bateaux de luxe » (*Tristes tropiques*, Paris : Plon, 1955, p. 15).

des langues vietiques, aux côtés de quelques autres collègues (parmi les phonéticiens : essentiellement Marc Brunelle et James Kirby), en contribuant à l'encadrement d'étudiants.

A la différence des parlers naish (naxi, na et lazé), que je parle avec plus ou moins d'aisance, je ne suis locuteur que d'une seule langue vietique : le vietnamien de Hanoi. Je connais néanmoins suffisamment bien la phonétique (synchronique et diachronique) du groupe vietique pour encadrer efficacement des recherches concernant ces diverses langues. J'ai en particulier dirigé le mémoire de Master 2 de Minh-Châu Nguyễn (2014-2016) au sujet du rôle de la glottalisation dans le système tonal du dialecte *mường* de Kim Thượng (province de Phú Thọ, Vietnam). Cette étude de phonétique expérimentale (Nguyễn 2016) a débouché sur un projet de thèse de doctorat, que je souhaiterais diriger.

Il va de soi que le choix de deux sous-groupes linguistiques (naish et vietique) comme domaine de spécialisation à long terme n'exclut pas de participer de façon ponctuelle à des programmes de recherche qui concerneraient d'autres langues : par exemple pour me rendre en février 2017 à l'invitation de collègues de Taïwan qui souhaitent discuter de questions de prosodie (intonation et tons) qui se posent dans la langue *saisyat* (famille austronésienne).

## **2.2. Perspective thématique : la « linguistique des langues » à l'ère numérique**

Au plan thématique, une de mes ambitions consiste à pratiquer une « Linguistique assistée par ordinateur », tirant parti des possibilités ouvertes par les technologies numériques. C'est à Jacqueline Vaissière que je dois l'impulsion initiale qui m'a lancé dans l'apprentissage du rudiment de la programmation, me conseillant d'apprendre à voler de mes propres ailes. Cela exige un investissement de temps conséquent, mais procure une liberté précieuse. Dès mes premiers pas dans Matlab et Praat, j'ai goûté la liberté promise. C'est ensuite sur la recommandation de Guillaume Jacques que je me suis initié à Perl (à partir de 2011), pour la manipulation d'expressions régulières. (Matlab ne gère pas Unicode, et n'est donc d'aucun secours pour les opérations sur les séquences de caractères phonétiques.) Je me suis également initié au système de composition de documents (typographie) LaTeX. Ces quelques bases en programmation ont facilité le dialogue avec les collègues chercheurs en informatique de l'Institut MICA (Hanoi, Vietnam), où j'ai séjourné

de novembre 2012 à juin 2016. La modélisation computationnelle demeure un horizon relativement distant : mes quelques réflexions dans ce domaine n'ont pas encore débouché sur des réalisations pratiques. Mais elles tiennent une part importante dans mes projets.

### **2.2.1. Tâches fondamentales pour la création de données connectées**

Un objectif à long terme que je poursuis, en collaboration avec des collègues du Lacito et d'autres laboratoires, consiste à expérimenter de nouveaux modèles de publication et d'exploitation des résultats de l'analyse linguistique. Il s'agit de constituer, pour une langue, un ensemble cohérent de ressources électroniques interconnectées – incluant grammaire, dictionnaire, corpus de textes, et documents multimédia. Ce travail peut contribuer à instaurer de nouvelles pratiques professionnelles dans le domaine de la description et la comparaison des langues, de nature à consolider l'assise empirique des formulations en typologie et théorie du langage.

Comme il a été mentionné en introduction, la linguistique de terrain vise traditionnellement à la production de grammaires, dictionnaires, et recueils de textes. Ces trois éléments forment ce que l'on appelle la « trilogie boasienne » (Foley 1999) par référence au travail fondateur de Franz Boas dans les premières années du XX<sup>e</sup> siècle (Boas 1902; Boas & Swanton 1911). Rien de computationnel dans cette méthode, formulée à une époque où les chercheurs publiaient leurs travaux sous forme imprimée. Mais un siècle plus tard, les technologies numériques ouvrent la possibilité de nouvelles avancées. Par l'ajout de la composante multimédia (enregistrements audio et vidéo), la trilogie est devenue *tétralogie* (Musgrave & Thieberger 2014). Ce tournant a été pris au Lacito dès 1994, par la fondation de la Collection Pangloss (Jacobson, Michailovsky & Lowe 2001; Michailovsky *et al.* 2014).

L'usage des nouvelles technologies va bien plus loin que la simple publication en ligne de travaux autrefois imprimés. Ce qui est désormais crucial, c'est la possibilité d'établir systématiquement des liens dynamiques entre les quatre volets de la tétralogie : dictionnaires, textes (et enregistrements multimédia) et grammaires électroniques peuvent être interconnectés l'un à l'autre. Les textes forment le cœur même des données linguistiques. Au fil de la lecture d'une grammaire sous forme papier, il est certes possible

de se reporter à des textes, et inversement. Mais en pratique, le lecteur est plus tenté d'aller consulter un exemple en contexte si celui-ci est à la portée d'un simple clic, que s'il est nécessaire de compulser un volume séparé. Les outils électroniques permettent donc une mise en résonance des diverses ressources concernant une même langue.

La création de liens hypertextes est, au fond, un simple prolongement du projet boasien de recueillir l'essentiel d'une langue et le rendre accessible. Plus que de *fixer* une langue au moyen de l'imprimé, il s'agit désormais de l'offrir à des modes nouveaux de navigation, en exploitant tout le potentiel de corpus en ligne, y compris par des traitements statistiques. L'annotation des documents multimédia est susceptible d'un perfectionnement indéfini ; et les généralisations linguistiques pourront bientôt s'élaborer dans une confrontation plus directe avec des observations empiriques et vérifiables, en dialogue avec les intuitions des locuteurs et des linguistes.

Le projet de grammaires électroniques n'est certes pas nouveau (Nordhoff 2008; Maxwell 2012), non plus que celui de ressources connectées. Mais en travaillant année après année à ce projet, les collègues au Lacito et dans des laboratoires aux problématiques proches seraient parmi les premiers à porter ces projets au stade de réalisations pratiques *de référence*, qui atteignent à un haut degré de précision et d'exhaustivité.

Cette composante de mes projets s'inscrit dans le sillage du projet ANR HimalCo (ANR-12-CORP-0006), coordonné par Guillaume Jacques. Ce projet a notamment permis la réalisation de dictionnaires japhug (Jacques 2015) et na (Michaud 2015b). Le dictionnaire japhug fait pour moi figure de modèle : il possède déjà 7 000 entrées, et offre une richesse bien supérieure à la moyenne des dictionnaires de « petites langues », y compris le dictionnaire na, qui n'en comprend à l'heure actuelle (2017) que 3 000. Une des raisons tient au fait que j'ai initialement donné la priorité au travail de phonologie et morphophonologie, sans consacrer au projet lexicographique l'attention nécessaire à sa croissance. L'enrichissement du dictionnaire, solidaire de la croissance du corpus de textes transcrits, est une entreprise de long terme, et comporte des difficultés indéniables : ainsi, l'exploration du vocabulaire religieux du na de Yongning nécessite une connaissance de la langue tibétaine, car les Na pratiquent le bouddhisme tibétain, dont ils ont emprunté le vocabulaire avec les croyances. Au cours des années qui viennent, je compte consacrer l'énergie nécessaire à des progrès réguliers dans l'enrichissement du dictionnaire. En parallèle, la mise en forme des données sera progressivement améliorée.

Une tâche fondamentale pour la suite du travail consistera à *lemmatiser* les textes : établir des liens hypertexte entre textes et dictionnaires, de façon à ce que chaque mot d'un texte pointe vers l'entrée de dictionnaire correspondante, et chaque entrée de dictionnaire donne accès à la liste intégrale des occurrences (concordance). Or les outils génériques de lemmatisation déjà existants n'ont qu'une efficacité très limitée dans le traitement de la morphotonologie, comme en langue na.

### 2.2.2. Vers une modélisation de haut niveau

Les règles morphologiques et morphophonologiques peuvent être implémentées dans un langage tel que Perl (tâche déjà entamée pour le na : création d'un générateur automatique des schémas tonals des combinaisons *numéral+classificateur*). Ces scripts permettent une confrontation systématique des données avec les règles proposées, d'où des avancées dans l'analyse. Des générateurs de paradigme permettent également d'inclure des paradigmes exhaustifs dans les dictionnaires. Néanmoins, du point de vue de la modélisation, Perl ne fait pas réellement partie des langages de haut niveau. Pour aller plus avant dans la modélisation, j'aimerais me familiariser avec les *transducteurs à états finis*, suivant l'exemple du traitement du yonaguni par Pellard & Yamada (sous presse). Pour traiter les effets à longue distance de la morphophonologie en langue na, un collègue informaticien (Mathieu Mangeot) me recommande d'essayer des langages spécialisés dans la programmation pour la linguistique (SLLP : *Specialized Languages for Linguistic Programming*).

Ce travail s'inscrit dans le contexte d'une Linguistique Assistée par Ordinateur. «Les Humanités Numériques vont dans le sens d'une approche dans laquelle une distinction nette entre humanités d'une part et informatique de l'autre n'est plus constructive» (Collins *et al.* 2015:10). S'il est clair que je suis linguiste et non informaticien, je m'efforce d'avoir une connaissance suffisante de l'informatique pour participer, dans le cadre de collaborations avec des ingénieurs et chercheurs en informatique, à la conception et au suivi du développement des outils logiciels. Ces outils ont vocation à faire partie de la «boîte à outils» du linguiste, car ils facilitent la confrontation des hypothèses avec les données, et permettent de parvenir à des implémentations logicielles des modèles linguistiques.

## Conclusion

L'enjeu d'un mémoire d'Habilitation à Diriger les Recherches est de faire ressortir – en l'inventant au besoin – la cohérence d'un parcours, pour inspirer confiance dans un projet d'avenir. Or mon parcours se construit en pratiquant (tant bien que mal) tout un éventail de disciplines bien différenciées (linguistique « de terrain », phonétique/phonologie, linguistique historique, traitement automatique des langues...), dont chacune est puissamment centrée sur sa propre gravité, et présente un dehors intimidant aux non-initiés. Ainsi, la phonétique expérimentale a des exigences de méthode qui lui sont propres, et qui tendent à faire de cette discipline un domaine quelque peu à part au sein des sciences du langage. Plusieurs linguistes du centre André-Georges Haudricourt m'ont témoigné de leur tracasserie lorsque telle de leurs observations en phonétique/phonologie était remise en cause (lors d'un exposé oral, ou de l'évaluation d'un manuscrit) au motif qu'elle ne reposait que sur de simples « impressions auditives » et n'avait pas reçu la sanction d'une étude expérimentale<sup>13</sup>. Or le souhait d'aller au-delà d'une description à l'oreille, et de faire appel à *ce qui se fait de mieux* en matière de phonétique expérimentale afin de parvenir à la « validation expérimentale », se heurte à des exigences aussi gourmandes en temps (d'apprentissage et de mise en œuvre) qu'exotiques au vu des pratiques courantes sur le terrain. Pour n'évoquer que la prise de données, il faut recruter des locuteurs en nombre suffisant – nombre qui connaît une certaine inflation au fil des ans : quatre locuteurs, cela suffisait jusque dans les années 2000, mais paraîtrait moins « représentatif » aujourd'hui – ; faire dire des mots dans des phrases-cadre afin de stabiliser leur réalisation phonétique ; soumettre à vérification, par des tests de perception, l'existence d'oppositions lexicales, pourtant démontrée à l'évidence par la pratique quotidienne de la langue... La complexité de l'entreprise tend à rebuter des linguistes « de terrain » dont l'emploi du temps est déjà bien rempli (voire chroniquement surchargé). De même, le processus de comparaison des langues et reconstruction pratiqué par le spécialiste de linguistique historique est une savante alchimie (mêlant érudition, expérience et intuition) qui a quelque chose de mystérieux pour le non-initié. Un collègue plus à l'aise sur la terre ferme (ou réputée telle) de la synchronie décrivait le flair du diachronicien expert comme

---

<sup>13</sup> Les méthodes non instrumentales sont regroupées sous le terme de « *impressionistic listening* », expression consacrée qui a quelque chose de dédaigneux, et gagnerait à mon avis à être remplacée par « *expert listening* » : l'enquêteur qui a fait l'effort d'entraîner son oreille et son gosier, et confronté ses intuitions à celles de collaborateurs dont c'est la langue maternelle, a acquis une expertise précieuse pour sa recherche.

une faculté « tératologique » : la formule facétieuse campe le diachronicien en figure aussi distante qu'elle est fascinante, personnage *bors du commun* et inimitable.

De la sorte, les forces centripètes propres à chaque discipline (du fait de la somme impressionnante de savoir et de savoir-faire qu'elle requiert) expliquent que ces disciplines puissent paraître mutuellement exclusives. Elles tendent à le devenir *de facto* dans la formation d'étudiants dont certains acquièrent un doctorat en sciences du langage sans avoir bénéficié d'une réelle formation en linguistique historique, et sans l'expérience du travail élémentaire qui consiste à établir le système phonologique d'une langue sur la base d'une enquête de première main.

Dans ces circonstances, il est problématique de dire que mon travail se situe au *croisement* des diverses disciplines évoquées dans les pages qui précèdent. Si leur intersection est vide, est-il recommandable de s'y établir ?

Une première réponse, sur un mode défensif – mais celui d'une défense de chacune des disciplines concernées –, est que mon parcours est *pluridisciplinaire* plutôt qu'interdisciplinaire. Je ne recherche pas l'hybridation entre disciplines. Au contraire : de la même manière que je m'efforce de pratiquer plusieurs langues sans les mélanger (séparation qui demande un effort constant), j'essaie d'éviter, dans mes travaux, la confusion des genres. Chaque spécialité (linguistique historique, phonétique/phonologie...) a ses méthodes, ses forces et sa logique. L'interdisciplinarité ne doit pas être le prétexte à des travaux de second rayon : être successivement diachronicien chez les expérimentalistes, et expérimentaliste chez les diachroniciens, pour s'exempter des exigences de l'une et l'autre discipline. Un travail de phonétique expérimentale doit être jugé à l'aune des critères de cette discipline. Il doit supporter la critique des spécialistes sans prétendre à un traitement de faveur au motif que l'auteur n'est qu'en partie du domaine. De même, l'analyse d'un système morphotonologique doit reposer sur les méthodes éprouvées de description linguistique ; s'il paraît utile de fournir quelques tracés de fréquence fondamentale, il doit être clair que c'est pour illustrer le propos, et non pour pallier les défaillances de l'argumentation linguistique au moyen de données phonétiques qui ne sont pas pertinentes à ce stade du raisonnement. Chacun de mes travaux doit être lisible par lui-même, délesté du contexte de l'auteur tel qu'il est raconté dans ces pages.

Cette mise au point effectuée, reste que je me trouve de fait à l'intersection de plusieurs domaines. La tradition orale (depuis longtemps fixée par l'écrit) de ma langue et



culture maternelle met en garde : « on ne peut être au four et au moulin » ; « qui trop embrasse mal étreint ». Mes lacunes dans tous mes domaines de spécialité sont bien réelles : pour ne citer que quelques-unes des plus criantes, au plan aréal je suis « tibéto-birmaniste » mais ignorant du tibétain aussi bien que du birman, et j'étudie des langues du groupe mon-khmer sans connaître ni le mon ni le khmer ; et au plan disciplinaire, je suis phonéticien mais n'ai plus suivi les nouvelles tendances en statistiques depuis dix ans. Conscient du problème, je m'efforce d'être vigilant : dans la direction des recherches d'étudiants, comme dans mes propres travaux, je me défie de mes opinions et premières impressions, et prends soin de compléter mon information avec l'aide de collègues.

## Références citées

---

- Amelot, Angélique & Alexis Michaud. 2006. Effets aérodynamiques du mouvement du velum : le cas des voyelles nasales du français. In *Actes des XXVII<sup>e</sup> Journées d'Etude de la Parole, Dinard (2006)*, 247–250. Dinard.
- Arvaniti, A., Robert Ladd & I. Mennen. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26. 3–25.
- Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.
- Blevins, Juliette. 2006. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32(2). 117–166.
- Boas, Franz. 1902. *Tsimshian texts*. (Bulletin of the Smithsonian Institution. Bureau of American Ethnology 27). Washington: Government Printing Office.
- Boas, Franz & James R. Swanton. 1911. Siouan (Dakota). *Handbook of American Indian Languages I*, 875–965. Washington: Government Printing Office, Bureau of American Ethnology, Bulletin 40.
- Bonin, Charles-Eudes. 1903. Vocabulaires. *T'oung Pao* series 2, 4. 117–126.
- Chahal, Dana & Sam Hellmuth. 2014. The intonation of Lebanese and Egyptian Arabic. In Sun-Ah Jun (ed.), *Prosodic typology II: the phonology of intonation and phrasing*, 365–404. Oxford: Oxford University Press.
- Chirkova, Katia & Yiya Chen. 2013a. Illustrations of the IPA: Lizu. *Journal of the International Phonetic Association* 43(1). 75–86.
- Chirkova, Katia & Yiya Chen. 2013b. Illustrations of the IPA: Xumi (part 1): Lower Xumi, the variety of the lower and middle reaches of the Shuiluo river. *Journal of the International Phonetic Association* 43(3). 363–379.
- Chirkova, Katia, Yiya Chen & Tanja Kocjančič Antolík. 2013. Illustrations of the IPA: Xumi (part 2): Upper Xumi, the variety of the upper reaches of the Shuiluo river. *Journal of the International Phonetic Association* 43(3). 381–396.
- Chirkova, Katia & Alexis Michaud. 2009. Approaching the prosodic system of Shixing. *Language and Linguistics* 10(3). 539–568.
- Chirkova, Katia, Dehe Wang, Yiya Chen, Angélique Amelot & Tanja Kocjančič Antolík. 2015. Illustrations of the IPA: Ersu. *Journal of the International Phonetic Association* 45(2). 187–211. doi:10.1017/S0025100314000437.
- Clements, Nick, Alexis Michaud & Cédric Patin. 2011. Do we need tone features? In Elizabeth Hume, John Goldsmith & W. Leo Wetzels (eds.), *Tones and Features*, 3–24. Berlin: De Gruyter Mouton.
- Collins, Sandra, Natalie Harrower, Dag Trygve Truslew Haug, Beat Immenhauser, Gerhard Lauer, Tito Orlandi, Laurent Romary & Eveline Wandl-Vogt. 2015. Going Digital: Creating Change in the Humanities. ALLEA.

- Degottex, Gilles, John Kane, Thomas Drugman, Tuomo Raitio & Stefan Scherer. 2014. COVAREP: a collaborative voice analysis repository for speech technologies. *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, 960–964. IEEE.
- Evans, Jonathan. 2008. “African” tone in the Sinosphere. *Language and Linguistics* 9(3). 463–490.
- Ferragne, Emmanuel & François Pellegrino. 2004. Rhythm in read British English: interdialect variability. *Proceedings of InterSpeech*.
- Ferragne, Emmanuel & François Pellegrino. 2008. Le rythme dans les dialectes de l’anglais: une affaire d’intensité? *Actes de Journées d’Etude de la Parole, Avignon*, 9–13. Avignon.
- Foley, William A. 1999. Review of Gerrit van Enk & Lourens de Vries, *The Korowai of Irian Jaya*. (Oxford studies in anthropological linguistics, 9). New York: Oxford University Press, 1997. Pp. xiv, 321. *Language in Society* 28(3). 470–472.
- François, Alexandre. 2003. *La sémantique du prédicat en mwotlap, Vanuatu*. Vol. 84. Louvain: Peeters.
- Gårding, Eva. 1998. Intonation in Swedish. In Daniel Hirst & Albert Di Cristo (eds.), *Intonation systems: A survey of twenty languages*, 112–130. Cambridge: Cambridge University Press.
- Gsell, René. 1979. Etudes et recherches tonales (Contribution à la typologie tonale). Paris: Université de la Sorbonne Nouvelle.
- Hagège, Claude & André-Georges Haudricourt. 1978. *La phonologie panchronique*. Paris: Presses Universitaires de France.
- Haudricourt, André-Georges. 1940. Méthode pour obtenir des lois concrètes en linguistique générale. *Bulletin de la Société de Linguistique de Paris* 41(1). 70–74.
- Haudricourt, André-Georges. 1956. De la restitution des initiales dans les langues monosyllabiques: le problème du thai commun. *Bulletin de la Société de Linguistique de Paris* 52. 307–322.
- Haudricourt, André-Georges. 1973. La linguistique panchronique nécessaire à la linguistique comparée, science auxiliaire de la diachronie sociologique et ethnographique. *Ethnies* 3. 23–26.
- Haudricourt, André-Georges. 1978. *A reconstruction of Neo-grammarians’ “sound laws” and the possibility of establishing “panchronic phonology.”* Tokyo: Project on Lexicological Analysis, National Inter-University Research Institute of Asian and African Languages and Cultures.
- Hé, Jírén 和即仁 & Zhúyí 姜竹仪 Jiāng. 1985. *Nàxīyǔ jiǎnzǐ 纳西语简志 (A brief description of the Naxi language)*. Beijing 北京: The Ethnic Publishing House 民族出版社.
- Hyman, Larry M. 2010. Kuki-Thaadow: an African tone system in Southeast Asia. In Floricic, Franck (ed.), *Essais de typologie et de linguistique générale. Mélanges offerts à Denis Creissels*, 31–51. Lyon: ENS Editions.
- Hyman, Larry M. & Kenneth VanBik. 2002. Tone and stem2 formation in Hakha Lai. *Linguistics of the Tibeto-Burman Area* 25. 113–121.

- Hyman, Larry M. & Kenneth VanBik. 2004. Directional rule application and output problems in Hakha Lai tone. *Language and Linguistics, Taipei: Academia Sinica, Special Issue: Phonetics and Phonology* 5. 821–861.
- Jacobson, Michel. 2004. Corpus oraux en linguistique de terrain. *Traitement automatique des langues* 45. 63–88.
- Jacobson, Michel, Boyd Michailovsky & John B. Lowe. 2001. Linguistic documents synchronizing sound and text. *Speech Communication* 33 [special issue: “Speech Annotation and Corpus Tools”]. 79–96.
- Jacques, Guillaume. 2014. Rapport en vue de l’habilitation à diriger les recherches. Paris: Institut National des Langues et Civilisations Orientales (INALCO).
- Jacques, Guillaume. 2015. *Dictionnaire japhug-chinois-français* 嘉绒 - 汉 - 法词典 . <https://hal.archives-ouvertes.fr/halshs-01244860v2/>.
- Jacques, Guillaume & Alexis Michaud. 2011. Approaching the historical phonology of three highly eroded Sino-Tibetan languages: Naxi, Na and Laze. *Diachronica* 28(4). 468–498.
- Kügler, Frank, Bernadett Smolibocki, Denis Arnold, Stefan Baumann, Bettina Braun, Martine Grice, Stefanie Jannedy, Jan Michalsky, Oliver Niebuhr & Jörg Peters. 2015. DIMA: Annotation guidelines for German intonation. *ICPhS 2015: 18th International Congress of Phonetic Sciences*. Glasgow: International Phonetic Association.
- Ladd, Robert. 2013. Tones and features: Phonetic and phonological perspectives ed. by John A. Goldsmith, Elizabeth Hume, and W. Leo Wetzels (review). *Language* 89(2). 360–363.
- Ladd, Robert, Dan Faulkner, Hanneke Faulkner & Astrid Schepman. 1999. Constant “segmental anchoring” of F0 movements under change in speech rate. *Journal of the Acoustical Society of America* 106. 1543–1554.
- Laniran, Yetunde O. 1992. Intonation in Tone Languages: the Phonetic Implementation of Tones in Yorùbá. Ithaca, New York: Cornell University.
- Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2016. *Languages of China: an Ethnologue country report*. ( 19th edition). Dallas: SIL International. <http://www.ethnologue.com/>.
- Lidz, Liberty. 2010. A descriptive grammar of Yongning Na (Mosuo). Austin: University of Texas, Department of linguistics. <https://repositories.lib.utexas.edu/bitstream/handle/2152/ETD-UT-2010-12-2643/LIDZ-DISSERTATION.pdf>.
- Mac, Dang-Khoa, Thi-Lan Nguyen, Alexis Michaud & Do-Dat Tran. 2015. Influences of speaker attitudes on glottalized tones: a study of two Vietnamese sentence-final particles. *Proceedings of ICPhS XVIII (18th International Congress of Phonetic Sciences)*. Glasgow.
- Martinet, André. 1996. *The internal conditioning of phonological systems*. Thiruvananthapuram: International School of Dravidian Linguistics.
- Maxwell, Mike. 2012. Electronic grammars and reproducible research. In Sebastian Nordhoff (ed.), *Electronic Grammaticography*, 207–235. (LD&C Special Publication 04). Honolulu: University of Hawaii Press.

- Mazaudon, Martine. 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, Martine & Alexis Michaud. 2005. Issues in the voice quality of Tamang tones: an electroglottographic study. *11th Himalayan Languages Symposium*. Chulalongkorn University, Bangkok.
- Mazaudon, Martine & Alexis Michaud. 2008. Tonal contrasts and initial consonants: a case study of Tamang, a “missing link” in tonogenesis. *Phonetica* 65(4). 231–256.
- Michailovsky, Boyd, Martine Mazaudon, Alexis Michaud, Séverine Guillaume, Alexandre François & Evangelia Adamou. 2014. Documenting and researching endangered languages: the Pangloss Collection. *Language Documentation and Conservation* 8. 119–135.
- Michailovsky, Boyd & Alexis Michaud. 2006. Syllabic inventory of a Western Naxi dialect, and correspondence with Joseph F. Rock's transcriptions. *Cahiers de linguistique - Asie Orientale* 35(1). 3–21.
- Michailovsky, Boyd, Alexis Michaud & Séverine Guillaume. 2011. A simple architecture for the fine-grained documentation of endangered languages: the LACITO multimedia archive. *Proceedings of Oriental-COCOSDA 2011, October 26th-28th, 2011*. Hsinchu, Taiwan.
- Michailovsky, Boyd, Alexis Michaud & He Xueguang. 2015. 骆克的纳西语-英语百科辞典中的拼音系统、对照国际音标. 茶马古道研究集刊 4. 119–124.
- Michaud, Alexis. 2002. Conservation des langues et partage des ressources: le rôle des chercheurs dans la mise en place de banques de données. *XXIV<sup>e</sup> Journées d'Etude de la Parole*, 153–156. Nancy, France.
- Michaud, Alexis. 2004. Final consonants and glottalization: new perspectives from Hanoi Vietnamese. *Phonetica* 61(2–3). 119–146.
- Michaud, Alexis. 2006a. Tonal reassociation and rising tonal contours in Naxi. *Linguistics of the Tibeto-Burman Area* 29(1). 61–94.
- Michaud, Alexis. 2006b. Replicating in Naxi (Tibeto-Burman) an experiment designed for Yorùbá: An approach to “prominence-sensitive prosody” vs. “calculated prosody.” *Speech Prosody 2006*, 819–822. Dresden.
- Michaud, Alexis. 2006c. Three extreme cases of neutralisation: nasality, retroflexion and lip-rounding in Naxi. *Cahiers de linguistique - Asie Orientale* 35(1). 23–55.
- Michaud, Alexis. 2008a. Phonemic and tonal analysis of Yongning Na. *Cahiers de linguistique - Asie Orientale* 37(2). 159–196.
- Michaud, Alexis. 2008b. Tones and intonation: some current challenges. *8th International Seminar on Speech Production (ISSP'08)*, 13–18. Strasbourg.
- Michaud, Alexis. 2009. Mùlǐ Shuǐtiánhuà shēngdiào xìtǒng yánjiū 木里水田话声调系统研究 (The prosodic system of Muli Shuitian (Laze)). *Minority Languages of China 民族语文* 6. 28–33.
- Michaud, Alexis. 2011a. The tones of numerals and numeral-plus-classifier phrases: on structural similarities between Naxi, Na and Laze. *Linguistics of the Tibeto-Burman Area* 34(1). 1–26.

- Michaud, Alexis. 2011b. Les systèmes de tons en Asie orientale : typologie, schémas évolutifs et modélisation. *Faits de Langues* 37. 247–261.
- Michaud, Alexis. 2012a. Yǒngníng Mósuōhuà (Nàyǔ) chángpiān yǔliào de jìlù zhěnglǐ yǔ yánjiù gōngzuò 永宁摩梭话（纳语）长篇语料的记录整理与研究（Ongoing documentation and research about the Yongning Na language (Mosuo)). *Lijiang Ethnic Studies* 丽江民族研究 5. 36–54.
- Michaud, Alexis. 2012b. The complex tones of East/Southeast Asian languages: current challenges for typology and modelling. *Keynote lecture at the Third International Symposium on Tonal Aspects of Languages (TAL 2012), Nanjing, China (2012)*. Nanjing.
- Michaud, Alexis. 2012c. Monosyllabicization: patterns of evolution in Asian languages. In Nicole Nau, Thomas Stolz & Cornelia Stroh (eds.), *Monosyllables: from phonology to typology*, 115–130. Berlin: Akademie Verlag. <http://halshs.archives-ouvertes.fr/halshs-00436432/>.
- Michaud, Alexis. 2013a. The tone patterns of numeral-plus-classifier phrases in Yongning Na: a synchronic description and analysis. In Nathan Hill & Tom Owen-Smith (eds.), *Transhimalayan Linguistics. Historical and Descriptive Linguistics of the Himalayan Area*, 275–311. (Trends in Linguistics. Studies and Monographs [TiLSM] 266). Berlin: De Gruyter Mouton.
- Michaud, Alexis. 2013b. Dān yīnjié huà: Dōngyà yǔyán zhōng de yīxiē chángjiàn yǎnhuà xíngshì 单音节化：东亚语言中的一些常见演化形式 (Monosyllabicization: patterns of evolution in East Asian languages). (Trans.) Lì Yī 衣莉 & Lín Yòujīng 林幼菁. *Chama Gudao Yanjiu Jikan* 茶马古道研究集刊 3. 74–96.
- Michaud, Alexis. 2013c. Studying level-tone systems in Asia: the case of the Naish languages. *Proceedings of International Conference on Phonetics of the Languages in China (ICPLC-2013)*, 1–6. Hong Kong.
- Michaud, Alexis. 2015a. Phrasing, prominence, and morphotonology: How utterances are divided into tone groups in Yongning Na. *Bulletin of Chinese Linguistics* 8. 86–116.
- Michaud, Alexis. 2015b. *Na-English-Chinese-French dictionary*. <https://halshs.archives-ouvertes.fr/halshs-01204638>.
- Michaud, Alexis. 2017. *Tone in Yongning Na: lexical tones and morphotonology*. (Studies in Diversity Linguistics). Berlin: Language Science Press.
- Michaud, Alexis & Marc Brunelle. 2016. Information structure in Asia: Yongning Na (Sino-Tibetan) and Vietnamese (Austroasiatic). In Caroline Féry & Shinichiro Ishihara (eds.), *Oxford Handbook of Information Structure*, 2048–2070. (Oxford Handbooks in Linguistics). Oxford: Oxford University Press.
- Michaud, Alexis, Michel Ferlus & Minh-Châu Nguyễn. 2015. Strata of standardization: the Phong Nha dialect of Vietnamese (Quảng Bình Province) in historical perspective. *Linguistics of the Tibeto-Burman Area* 38(1). 124–162.
- Michaud, Alexis, Séverine Guillaume, Guillaume Jacques, Đăng-Khoa Mạc, Michel Jacobson, Thu Hà Phạm & Matthew Deo. 2016. Contribuer au progrès solidaire des recherches et de la documentation : la Collection Pangloss et la Collection AuCo. *Actes de la conférence conjointe JEP-TALN-RECITAL 2016, volume 1 : Journées d'Etude de la Parole*, vol. 1, 155–163. Paris: Association Francophone de la Communication Parlée.

- Michaud, Alexis, Andrew Hardie, Séverine Guillaume & Martine Toda. 2012. Combining documentation and research: Ongoing work on an endangered language. In Xiong Deyi, Eric Castelli, Dong Minghui & Pham Thi Ngoc Yen, (eds.), *Proceedings of LALP 2012 (2012 International Conference on Asian Language Processing)*, 169–172. Hanoi, Vietnam: MICA Institute, Hanoi University of Science and Technology.
- Michaud, Alexis & Likun He. 2015. Phonemic and tonal analysis of the Pianding dialect of Naxi (Dadong County, Lijiang Municipality). *Cahiers de Linguistique - Asie Orientale* 44(1). 1–35 plus online Appendices.
- Michaud, Alexis & Xueguang He. 2007. Reassociated tones and coalescent syllables in Naxi (Tibeto-Burman). *Journal of the International Phonetic Association* 37(3). 237–255.
- Michaud, Alexis & Xueguang He. 2010. Lǐjiāngshì Gǔchéngqū Wénhuàxiāng Nàxīyǔ yīnwèi xìtǒng yánjiū 丽江市古城区文化乡纳西语音位系统研究 (The phonemic system of A-sher (Wenhua) Naxi). *Chama Gudao Yanjiu Jikan* 茶马古道研究集刊 1. 190–202.
- Michaud, Alexis & Guillaume Jacques. 2010. Insights into Naxi and Pumi at the end of the 19th century: evidence on sound changes from the word lists by Charles-Eudes Bonin. *Cahiers de Linguistique - Asie Orientale* 39(1). 21–40.
- Michaud, Alexis & Guillaume Jacques. 2012. The phonology of Laze: phonemic analysis, syllabic inventory, and a short word list. *Yuyanxue Luncong* 语言学论丛 45. 196–230.
- Michaud, Alexis, Guillaume Jacques & Robert L. Rankin. 2012. Historical transfer of nasality between consonantal onset and vowel: from C to V or from V to C? *Diachronica* 29(2). 201–230.
- Michaud, Alexis, Guillaume Jacques & Robert L. Rankin. 2014. 鼻化特征在历史演变中的转移——从辅音声母到元音？还是元音到辅音声母？. (Trans.) Li Yi 衣莉. *EastLing* (《东方语言学》) 14. 159–188.
- Michaud, Alexis & Barbara Kühnert. A paraître. A pilot study on the F0 curve of syllable-initial sonorants, comparing nasals, lenis stops and fortis stops. In Nicolas Ballier (ed.), *Actes du Congrès 2006 de l'ALAES*. Paris.
- Michaud, Alexis & Barbara Kühnert. 2006. La courbe de F0 des sonantes initiales de syllabe joue-t-elle un rôle prosodique? Etude-pilote de données d'anglais britannique. *XXVIIe Journées d'Etude de la Parole*, 121–124. Dinard.
- Michaud, Alexis & Latami Dashi. 2011. A description of endangered phonemic oppositions in Mosuo (Yongning Na). In Tjeerd De Graaf, Xu Shixuan & Cecilia Brassett (eds.), *Issues of language endangerment*, 55–71. Beijing 北京: Intellectual Property Publishing House 知识产权出版社.
- Michaud, Alexis & Lātāmī Dáshí 拉他咪·达石. 2010. Yúnnánshěng Lǐjiāngshì Yǒngníng qūyù Mósuōhuà zhōng bīnlín xiāoshī de shēngdiào yǔ yīnwèi 云南省丽江市永宁区域摩梭话中濒临消失的声调与音位 (Endangered tones and phonemes in Yongning Na). *Lijiang Ethnic Studies* 丽江民族研究 4. 344–355.
- Michaud, Alexis & Martine Mazaudon. 2006. Pitch and voice quality characteristics of the lexical word-tones of Tamang, as compared with level tones (Naxi data) and pitch-plus-voice-quality tones (Vietnamese data). *Speech Prosody 2006*, 823–826. Dresden.

- Michaud, Alexis & Jacqueline Vaissière. 2007a. Le devenir phonétique des formes rédupliquées : réduplication, tons lexicaux et intonation en naxi. *Faits de langues* 29. 23–35.
- Michaud, Alexis & Jacqueline Vaissière. 2007b. The phonetic evolution of reduplicated expressions: reduplication, lexical tones and prosody in Na (Naxi). In Jürgen Trouvain & William Barry (eds.), *16th International Congress of Phonetic Sciences*, 801–804. Saarbrücken.
- Michaud, Alexis & Jacqueline Vaissière. 2008. Reflections about the phonemic analysis of Yongning Na (Tibeto-Burman): Perceptual transcription and acoustic data. 第八届中国语音学学术会议 (*8th Phonetics Conference of China*) (Full-paper proceedings published on CD-ROM, without page numbers). Beijing, April 18th–20th 2008.
- Michaud, Alexis & Jacqueline Vaissière. 2009. Perceptual transcription and acoustic data: the example of /i/ in Yongning Na (Tibeto-Burman). *Chinese Journal of Phonetics* (中国语音学报) 2. 10–17.
- Michaud, Alexis & Jacqueline Vaissière. 2015. Tone and intonation: introductory notes and practical recommendations. *KALIPHO - Kieler Arbeiten zur Linguistik und Phonetik* 3. 43–80.
- Michaud, Alexis, Jacqueline Vaissière & Minh-Châu Nguyễn. 2015. Phonetic insights into a simple level-tone system: “careful” vs. “impatient” realizations of Naxi High, Mid and Low tones. *Proceedings of ICPbS XVIII (18th International Congress of Phonetic Sciences)*. Glasgow.
- Michaud, Alexis & Tuân Vu-Ngoc. 2004. Glottalized and nonglottalized tones under emphasis: open quotient curves remain stable, F0 curve is modified. In Bernard Bel & Isabelle Marlien (eds.), *Speech Prosody 2004*, 745–748. Nara, Japan.
- Michaud, Alexis & Tuân Vu-Ngoc. 2007. Các thanh điệu thanh hầu hoá và phi thanh hầu hoá dưới tác dụng của sự nhấn âm: các giá trị hệ số mở tiến dần đến cực đại. *Ngôn ngữ* 213(2). 46–52. (Translation of the preceding reference: Michaud, Alexis & Tuân Vu-Ngoc 2004.)
- Michaud, Alexis, Tuân Vu-Ngoc, Angélique Amelot & Bernard Roubeau. 2006. Nasal release, nasal finals and tonal contrasts in Hanoi Vietnamese: an aerodynamic experiment. *Mon-Khmer Studies* 36. 121–137.
- Michaud, Alexis & Xu Jirong 徐继荣. 2012. Xiānggélǎ xiàn Ciēndīng cūn nàxīyǔ yīnxì yánjiū 香格里拉县次恩丁村纳西语音系研究 (Research about the phonemic system of the Naxi dialect of Ciending village, Shangri-la prefecture). *Chama Gudao Yanjiu Jikan* 茶马古道研究集刊 2. 139–165.
- Musgrave, Simon & Nicholas Thieberger. 2014. Rethinking grammatical description: from Heath to hypertext. Lecture. Research Unit for Indigenous Language, University of Melbourne. <https://indiglang.arts.unimelb.edu.au/events/rethinking-grammatical-description-from-heath-to-hypertext/>.
- Nguyễn, Minh-Châu. 2016. The tone system of Kim Thượng Mường: an experimental study of fundamental frequency, duration, and phonation types. Hanoi: VNU-USSH - Vietnam National University - Department of Linguistics M.A. thesis. <https://dumas.ccsd.cnrs.fr/dumas-01405496/>.



- Nguyen, Thi-Lan, Alexis Michaud, Do-Dat Tran & Dang-Khoa Mac. 2013. The interplay of intonation and complex lexical tones: how speaker attitudes affect the realization of glottalization on Vietnamese sentence-final particles. *Proceedings of Interspeech 2013*. Lyon.
- Niebuhr, Oliver & Alexis Michaud. 2015. Speech data acquisition: the underestimated challenge. *KALIPHO - Kieler Arbeiten zur Linguistik und Phonetik* 3. 1–42.
- Nordhoff, Sebastian. 2008. Electronic reference grammars for typology: challenges and solutions. *Language Documentation and Conservation* 2(2). 296–324.
- Öhman, Sven & Jan Lindqvist. 1968. Analysis-by-synthesis of prosodic pitch contours. *STUF-Language Typology and Universals* 21(1–6). 164–170.
- Pain, Frédéric, Michel Ferlus, Alexis Michaud & Thu Hà Phạm. 2014. *EFEO-CNRS-SOAS word list for linguistic fieldwork in Southeast Asia*. Hanoi: International Research Institute MICA. <https://halshs.archives-ouvertes.fr/halshs-01068533/>.
- Pellard, Thomas & Masahiro Yamada. In press. Verb morphology and conjugation classes in Dunan (Yonaguni). In Ferenc Kiefer, James P. Blevins & Huba Bartos (eds.), *Morphological paradigms and functions*. Leiden: Brill.
- Pellegrino, François, Egidio Marsico, Ioana Chitoran & Christophe Coupé. 2009. *Approaches to phonological complexity*. Berlin/New York: Walter de Gruyter.
- Post, Brechtje, Emmanuel A. Stamatakis, Iwo Bohr, Francis Nolan & Chris Cummins. 2015. Categories and gradient in intonation: a functional Magnetic Resonance Imaging study. *The Phonetics–Phonology Interface: Representations and methodologies*, vol. 335, 259–284. Amsterdam: John Benjamins.
- Prieto, P., Jan van Santen & J. Hirschberg. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* 23. 429–451.
- Prieto, Pilar. 2009. Tonal alignment patterns in Catalan nuclear falls. *Lingua* 119(6). 865–880.
- Rouas, Jean-Luc, Jérôme Farinas, François Pellegrino & Régine André-Obrecht. 2005. Rhythmic unit extraction and modelling for automatic language identification. *Speech Communication* 47(4). 436–456.
- Sagart, Laurent. 1993. *Les dialectes gan. Etudes sur la phonologie et le lexique d'un groupe de dialectes chinois*. Paris: Langages croisés.
- Silverman, Kim, Mary Beckman, J. Pitrelli, M. Ostendorf, Colin W. Wightman, P. Price, Janet Pierrehumbert & J. Hirschberg. 1992. ToBI: A Standard for Labeling English Prosody. *Proceedings of the 1992 International Conference on Spoken Language Processing*, vol. 2, 867–870. Banff, Canada.
- Vaissière, Jacqueline. 2002. Cross-linguistic prosodic transcription: French vs. English. In N.B. Volskaya, N.D. Svetozarova & P.A. Skrelin (eds.), *Problems and methods of experimental phonetics. In honour of the 70th anniversary of Pr. L.V. Bondarko*, 147–164. St Petersburg: St Petersburg State University Press.
- Vaissière, Jacqueline. 2004. The perception of intonation. In David B. Pisoni & Robert E. Remez (eds.), *Handbook of Speech Perception*, 236–263. (Blackwell Textbooks in Linguistics). Oxford, U.K. & Cambridge, Massachusetts: Blackwell.

- Vaissière, Jacqueline & Alexis Michaud. 2006. Prosodic constituents in French: a data-driven approach. In Ivan Fónagy, Yuji Kawaguchi & Tsunekazu Moriguchi (eds.), *Prosody and syntax: Cross-linguistic perspectives*, 47–64. (Usage-Based Linguistic Informatics). Amsterdam: John Benjamins.
- Vu-Ngoc Tuân, Christophe d’Alessandro & Alexis Michaud. 2005. Using open quotient for the characterization of Vietnamese glottalized tones. *Eurospeech-Interspeech 2005: 9th European Conference on Speech Communication and Technology*, 2885–2889. Lisboa.
- Xu, Yi & Emily Q. Wang. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33. 319–337.
- Zhíbā, Ěrchē 直巴·尔车 & Ruijuān Xǔ 许瑞娟. 2013. *Mósuōyǔ chángyòng cíjù huìcù* 摩梭语常用词句荟萃 (*An anthology of everyday words and expressions in the Mosuo language*). Kunming 昆明: 云南人民出版社.

# Table des matières

---

Introduction .....	5
<b>1. Travaux réalisés.....</b>	<b>10</b>
<b>1.1. Linguistique de terrain.....</b>	<b>10</b>
1.1.1. Etude de la langue na de Yongning .....	11
1.1.2. Etude d'autres parlers du groupe naish .....	16
<b>1.2. Phonétique expérimentale et phonologie théorique .....</b>	<b>21</b>
<b>1.2.1. Phonétique expérimentale .....</b>	<b>21</b>
1.2.1.1. <i>Une étude au sujet de la réduction syllabique en langue naxi .....</i>	<i>22</i>
1.2.1.2. <i>Etude de la langue tamang.....</i>	<i>23</i>
1.2.1.3. <i>Etude expérimentale de la nasalité .....</i>	<i>25</i>
<b>1.2.2. Phonologie théorique.....</b>	<b>25</b>
<b>1.3. Prosodie : études de cas et réflexions générales.....</b>	<b>27</b>
1.3.1. Etudes de cas .....	27
1.3.2. Réflexions générales et typologie prosodique .....	28
<b>1.4. Phonétique historique .....</b>	<b>31</b>
1.4.1. Etude de la phonologie historique des langues naish.....	32
1.4.2. Etude du transfert de la nasalité entre un groupe de consonnes initial et une voyelle .....	33
1.4.3. Autres travaux en phonétique historique .....	34
<b>1.5. Méthodologie : de la collecte à l'archivage des données   linguistiques .....</b>	<b>36</b>
<b>1.5.1. Méthodes de collecte et analyse de données.....</b>	<b>37</b>
1.5.1.1. <i>Principes.....</i>	<i>37</i>
1.5.1.2. <i>Un outil logiciel : script pour l'analyse du signal électroglottographique.....</i>	<i>39</i>
1.5.1.3. <i>Un questionnaire linguistique électronique.....</i>	<i>41</i>
<b>1.5.2. Archivage et diffusion des données .....</b>	<b>43</b>
1.5.2.1. <i>Principes.....</i>	<i>43</i>
1.5.2.2. <i>Principale réalisation : numérisation des données du linguiste Michel Ferlus.....</i>	<i>46</i>

<b>2. Orientations futures</b> .....	<b>53</b>
2.1. Perspective aréale .....	54
2.1.1. Domaine des langues naish (sino-tibétain).....	54
2.1.2. Domaine des langues vietiques (austro-asiatique) .....	57
2.2. Perspective thématique : la « linguistique des langues » à l'ère numérique.....	58
2.2.1. Tâches fondamentales pour la création de données connectées .....	59
2.2.2. Vers une modélisation de haut niveau .....	61
<b>Conclusion</b> .....	<b>62</b>
<b>Références citées</b> .....	<b>65</b>
<b>Table des matières</b> .....	<b>74</b>