Pataquès et liaison : étude de deux phénomènes de sandhi externe dans des corpus de français chanté

Gasparde Coutanson

▶ **To cite this version:**

Gasparde Coutanson. Pataquès et liaison : étude de deux phénomènes de sandhi externe dans des corpus de français chanté. Linguistique. Université de Nanterre - Paris X, 2023. Français. NNT : 2023PA100074 . tel-04404078

HAL Id: tel-04404078
https://theses.hal.science/tel-04404078

Submitted on 18 Jan 2024

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.

Université Paris Nanterre

200 av. de la République
92001 Nanterre Cedex
www.parisnanterre.fr

Membre de l'université Paris Lumières

École doctorale 139 :
Connaissance, langage, modélisation
Laboratoire MoDyCo - UMR 7114

Gasparde Coutanson

Pataquès et liaison : étude de deux phénomènes de sandhi externe dans des corpus de français chanté

Thèse présentée et soutenue publiquement le 02/10/2023
en vue de l'obtention du doctorat de Sciences du langage
de l'Université Paris Nanterre
sous la direction de M. Olivier Baude (Université Paris Nanterre)

Jury :

Rapporteure :	Elissa Pustka	Professeure, Université de Vienne
Rapporteur :	Gabriel Bergounioux	Professeur émérite, Université d'Orléans
Membre du jury :	Céline Dugua	Maitre de conférences, Université d'Orléans
Membre du jury :	Sophie Wauquier	Professeure, Université Paris 8 Vincennes-St Denis
Membre du jury :	David Hornsby	Professeur émérite, Université du Kent
Membre du jury :	Christophe Parisse	Chargé de recherche INSERM, HDR, Université Paris Nanterre
Directeur :	Olivier Baude	Professeur, Université Paris Nanterre

Pataquès et liaison : étude de deux phénomènes de sandhi externe dans des corpus de français chanté

Gasparde Coutanson

Thèse présentée et soutenue publiquement le 2 octobre 2023
en vue de l'obtention du doctorat de Sciences du langage
de l'Université Paris Nanterre
sous la direction de M. Olivier Baude

Préparée au sein du laboratoire Modèles, Dynamiques, Corpus (MoDyCo) - UMR 7114
École doctorale Connaissance, langage, modélisation (ED 139)
Université Paris Nanterre

Jury :

Elissa PUSTKA, Professeure, Université de Vienne (Rapporteure)
Gabriel BERGOUNIOUX, Professeur émérite, Université d'Orléans (Rapporteur)
Céline DUGUA, Maitre de conférences, Université d'Orléans (Membre du jury)
Sophie WAUQUIER, Professeure, Université Paris 8 Vincennes-St Denis (Membre du jury)
David HORNSBY, Professeur émérite, Université du Kent (Membre du jury)
Christophe PARISSE, Chargé de recherche INSERM, HDR, Université Paris Nanterre (Membre du jury)
Olivier BAUDE, Professeur, Université Paris Nanterre (Directeur de thèse)

Résumé

Dans le cadre de cette thèse doctorale, je m'intéresse au pataquès, un phénomène morphophonologique rare se produisant notamment en français. Apparenté à la liaison, autre phénomène qui suscite mon intérêt, ce sandhi externe inattendu constitue un trait d'oralité généralement non orthographié. Cette recherche m'a amenée à m'interroger sur la méthodologie à adopter en linguistique pour étudier des phénomènes rares et dépasser les collections d'exemples. Je présente aussi deux corpus transcrits et annotés pour l'occasion : un corpus de tops 1, comprenant des chansons de l'industrie du disque francophones à succès, ayant atteint la première place de classements musicaux français entre 1956 et 2017, ainsi que le fonds Jean Dumas, une collection de chants de tradition orale recueillis dans le centre de la France dans les années 1960. Mon travail repose par ailleurs sur divers corpus finalisés (corpus acquisitionnel, corpus variationniste, atlas linguistique…). Mes deux corpus principaux me permettent de réaliser une analyse qualitative et quantitative de la liaison dans la langue chantée, ainsi que du phénomène de pataquès, encore peu étudié, mais aussi une analyse micro-diachronique de la liaison dans un nouveau corpus de professionnels de la parole.

Mots-clés

morphophonologie
sandhi externe
liaison
pataquès
phénomènes rares
linguistique sur corpus
chansons de tradition orale

Abstract

***Pataquès* and *liaison*: a study of two external sandhi phenomena in corpora of sung French**

This doctoral thesis focuses on a low-frequency linguistic event, *pataquès*, a morphophonological phenomenon almost never written. This unexpected case of external sandhi is related to the more well-known phenomenon of *liaison*, and both occur in French, among other languages. This research considers a methodology to work on rare linguistic events and go beyond example collections. I also propose two newly transcribed and annotated corpora: a corpus of number-one singles, which contains songs from the music industry that were successful in France between 1956 and 2017, and the Jean Dumas collection, which includes folk songs recorded in the 1960s in central France. This study also relies on different types of corpora (parent-child interactions, variationist data, a linguistic atlas, etc.). The two main corpora allow a qualitative and quantitative study of *liaison* in sung corpora, but also of *pataquès*, as this phenomenon remains largely understudied. Finally, I can carry out a micro-diachronic study of *liaison* among new professional speakers.

Keywords

morphophonology
external sandhi
liaison
pataquès
rare events
corpus linguistics
folk songs

Table des matières

Table des figures **10**

Liste des tableaux **12**

Remerciements **14**

Introduction **16**

1 Sandhi, liaison et pataquès : d'hyperonyme en hyponyme **19**
1.1 La liaison : un cas de sandhi externe 19
1.2 La liaison : un phénomène complexe 21
1.2.1 La liaison : définition 21
1.2.2 La liaison : facteurs linguistiques 22
1.2.2.1 Liaison et phonologie 22
1.2.2.1.1 Liaison et hiatus 23
1.2.2.1.2 Liaison et syllabation 26
1.2.2.1.3 Nature et prégnance des consonnes de liaison 28
1.2.2.1.3.1 Nature des consonnes de liaison au niveau phonologique 28
1.2.2.1.3.2 Nature des consonnes de liaison au niveau phonétique 30
1.2.2.1.3.3 Prégnance des consonnes de liaison . . 32
1.2.2.1.4 Liaison et segments environnants 35
1.2.2.2 Liaison et morphologie 38
1.2.2.2.1 Statut morphologique des consonnes de liaison 38
1.2.2.2.2 Liaison et paradigmes morphologiques 39
1.2.2.2.3 Morphologisation des consonnes de liaison . . 40
1.2.2.2.4 Liaison et neutralisation morphologique 42

1.2.2.3 Liaison et syntaxe 44
1.2.2.3.1 Liaison et cohésion syntaxique 44
1.2.2.3.2 Typologie des contextes de liaison : nommer et définir les types retenus 46
1.2.2.3.3 Typologie des contextes de liaison : évolution des contextes retenus pour chaque type 48
1.2.2.4 Liaison et lexique 52
1.2.2.5 Liaison et sémantique 54
1.2.3 La liaison : facteurs sociolinguistiques 55
1.2.3.1 Liaison et variation diachronique 55
1.2.3.1.1 La liaison en phonétique diachronique 55
1.2.3.1.1.1 Données en phonétique diachronique . 55
1.2.3.1.1.2 Évolution diachronique de la liaison : dynamiques contradictoires 56
1.2.3.1.1.3 Sort des consonnes finales et de quelques segments initiaux spécifiques 60
1.2.3.1.1.4 La liaison en français contemporain : corpus oraux microdiachroniques 65
1.2.3.1.2 Variation diachronique, liaison et morphologie 71
1.2.3.1.2.1 Statut morphologique des consonnes de liaison en diachronie 71
1.2.3.1.2.2 Origine des consonnes flexionnelles . . 72
1.2.3.2 Liaison et variation diamésique 77
1.2.3.2.1 Liaison et graphie 77
1.2.3.3 Liaison et variation diatopique 81
1.2.3.4 Liaison et variation diaphasique 82
1.2.3.4.1 Liaison et lecture 82
1.2.3.4.2 Liaison et versification 84
1.2.3.4.3 Liaison et chanson 86
1.2.4 Liaison et variation diastratique 87
1.2.4.1 Liaison et différences sociolectales 87
1.2.4.2 Liaison et différences de niveau d'étude 88
1.3 Le pataquès : un cas particulier de liaison 90
1.3.1 Le pataquès : à la recherche d'une appellation 90
1.3.2 Le pataquès : à la recherche d'une définition 91
1.3.2.1 Le pataquès : une liaison non normative ? 91

1.3.2.2 Le pataquès : une liaison non standard ? 93
1.3.2.3 Le pataquès : une liaison non étymologique, non historique ? 94
1.3.2.4 Le pataquès : une liaison inattendue 96
1.3.3 Autres productions erratiques en matière de liaison 98
1.3.4 Sous-catégories de pataquès 100
1.3.5 Le pataquès : facteurs linguistiques 102
1.3.5.1 Pataquès et phonologie 102
1.3.5.1.1 Nature et prégnance des consonnes de pataquès 102
1.3.5.1.2 Pataquès et segments environnants 104
1.3.5.1.3 Pataquès en //ʁ// : marque pseudo-morphologique déclenchée par un contexte phonétique similaire 104
1.3.5.2 Pataquès et morphologie : morphologisation des consonnes de liaison 105
1.3.5.3 Pataquès et syntaxe : contextes syntaxiques d'apparition de pataquès connus 107
1.3.5.4 Pataquès et sémantique 111
1.3.6 Le pataquès : facteurs sociolinguistiques 112
1.3.6.1 Pataquès et variation diachronique 112
1.3.6.2 Pataquès et variation diamésique 114
1.3.6.2.1 Pataquès et graphie 114
1.3.6.3 Pataquès et variation diatopique 117
1.3.6.4 Pataquès et variation diagénique 119

2 Méthodologie **121**
2.1 Méthodologie générale 121
2.1.1 État de l'art et collection d'exemples attestés 121
2.1.2 Choix et carottage de corpus disponibles variés 125
2.1.3 Constitution de nouveaux corpus 132
2.1.3.1 Obtention d'un fichier son et mise en place d'un plan de nommage 133
2.1.3.2 Préparation d'un fichier .cha 134
2.1.3.3 Alignement son/texte 136
2.1.3.4 Transcription et annotation du document sonore 137
2.1.3.4.1 Transcription et annotation sous CLAN 137
2.1.3.4.2 Annotation de la liaison et du pataquès 140

2.1.3.5 Réécoute et vérification 150
2.1.4 Analyse fine des corpus les plus pertinents 151
2.2 Traitement du corpus de tops 1 152
2.2.1 Téléchargement du fichier son et mise en place d'un plan de nommage 152
2.2.2 Préparation d'un fichier .cha 156
2.2.3 Obtention des paroles et alignement son/texte 156
2.2.4 Transcription et annotation du document sonore 157
2.2.5 Réécoute et vérification 158
2.3 Traitement du fonds Jean Dumas 158
2.3.1 Téléchargement du fichier son et mise en place d'un plan de nommage 159
2.3.2 Préparation d'un fichier .cha 161
2.3.3 Alignement son/texte 163
2.3.4 Transcription et annotation du document sonore 164
2.3.5 Réécoute et vérification 174

3 Présentation du fonds Jean Dumas et du corpus de tops 1 177
3.1 Ferdinand Brunot et les Archives de la Parole : premiers enregistrements de la langue française 177
3.2 Fonds Jean Dumas 178
3.2.1 Choix d'un corpus de chansons de tradition orale 178
3.2.2 Jean Dumas, folkloriste à l'origine du fonds Jean Dumas 182
3.2.3 Démarche de Jean Dumas 185
3.2.3.1 Forces de la démarche de Jean Dumas 185
3.2.3.2 Faiblesses de la démarche de Jean Dumas 192
3.3 Corpus de tops 1 197

4 Résultats et discussion : carottage et première fouille des corpus 203
4.1 Carottage des corpus choisis 203
4.1.1 Présentation des corpus carottés 204
4.1.1.1 Fonds Jean Dumas 204
4.1.1.2 Corpus Phonologie du Français Contemporain 205
4.1.1.3 Corpus Acquisition de la Liaison et Interaction Parents Enfants 207
4.1.1.4 Atlas Linguistique du Francoprovençal Valaisan 208

4.1.1.5 Fonds Charles Joisten . . . 209
4.1.2 Méthodologie de carottage, voire de première fouille des données 210
4.1.3 Résultats du carottage . . . 212
4.1.3.1 Résultats du carottage de l'ALAVAL . . . 213
4.1.3.2 Résultats du carottage du corpus ALIPE . . . 218
4.1.3.3 Résultats du carottage du FJD . . . 219
4.1.3.4 Résultats du carottage du fonds Charles Joisten . . . 220
4.1.3.5 Résultats du carottage du corpus PFC . . . 223
4.1.4 Résultats de la première fouille des corpus . . . 226
4.1.4.1 Résultats de la première fouille du corpus ALIPE . . . 226
4.1.4.2 Résultats de la première fouille du corpus PFC . . . 240

5 Résultats et discussion : fouille du FJD et du corpus de tops 1 248
5.1 Dénasalisations surprenantes liées au phénomène de liaison dans le FJD 248
5.1.1 Dénasalisation de la voyelle finale du Mot 1 (M1) hors contexte de liaison . . . 248
5.1.2 Fermeture et dénasalisation de la voyelle de *un* en contexte de liaison . . . 249
5.2 Productions erratiques en matière de liaisons dans le FJD et le corpus de tops 1 . . . 251
5.2.1 Non-liaisons attestées surprenantes dans le FJD et le corpus de tops 1 . . . 251
5.2.2 Liaisons en contexte disjonctif dans le FJD . . . 253
5.2.3 Liaisons attestées surprenantes dans le FJD et le corpus de tops 1 255
5.3 Liaison et lexique : actualisation des connaissances à l'aide des données du FJD et du corpus de tops 1 . . . 259
5.3.1 Adjectifs masculins prénominaux en contexte de liaison dans le FJD et le corpus de tops 1 . . . 259
5.3.2 Mots 1 liaisonnants pour les consonnes de liaison les moins fréquentes dans le FJD et le corpus de tops 1 . . . 261
5.4 Liaison et pataquès dans deux corpus chantés . . . 262
5.4.1 Liaisons et liaisons variables dans le FJD et le corpus de tops 1 . 262
5.4.2 La liaison en microdiachronie dans le corpus de tops 1 . . . 264
5.4.3 Pataquès dans le FJD et le corpus de tops 1 . . . 266
5.4.4 Ordonnancement des consonnes de liaison et de pataquès dans le FJD . . . 268

5.5 Lien entre les phénomènes de pataquès et de liaison dans le FJD et le corpus de tops 1 270
5.5.1 Origine des pataquès dans la construction NOM singulier + *et* pour le FJD 270
5.5.2 Origine des pataquès dans la construction X + *ensemble* pour le FJD 272
5.5.3 Origine des pataquès dans la construction VER conjugué + X pour le FJD et le corpus de tops 1 274

Conclusion **278**

Annexes **300**
I Références des énoncés sélectionnés lors du carottage de l'ALAVAL . . . 300
II Requêtes utilisées lors de la première fouille du corpus ALIPE 311
III Autres résultats aux requêtes utilisées lors de la première fouille du corpus ALIPE 313
IV Requêtes utilisées lors de la première fouille du corpus PFC 323
V Liaisons codées comme épenthétiques dans le corpus PFC 325

Abréviations **334**

Acronymes **336**

Index **339**

Crédits des illustrations **342**

Bibliographie **344**
Bibliographie scientifique 344
Cours universitaires 362
Corpus 362
Programmes, logiciels et outils informatiques, manuels 362
Dictionnaires, entrées de dictionnaires 364
Sitographie 365
Autres références 367

Table des figures

1.1 Élision du mot *hiatus* dans le corpus French 2019 (Google Ngram Viewer) 23
1.2 Schéma de la structure syllabique 26
1.3 Consonnes de liaison du français et autres consonnes impliquées dans des sandhis consonantiques 28
1.4 Part de réalisation de chaque consonne de liaison ou de pataquès dans l'ensemble des liaisons ou pataquès réalisés 34
1.5 Part de réalisation potentielle de chaque consonne de liaison dans l'ensemble des contextes de liaison 35
1.6 Part de réalisation de chaque consonne de liaison dans l'ensemble des réalisations potentielles de cette même liaison 35
1.7 Tableau simplifié de la classification de Delattre (1947 : 152) 49
1.8 Nasalisation en français : conservation de la voyelle nasale ou de la consonne nasale aux XVI^{ème}-XVII^{ème} siècle (s.) 61
1.9 Évolution des taux de liaisons variables réalisées entre les deux périodes d'enregistrement chez les sept locuteurs du module Diachronie dans le corpus des Enquêtes sociolinguistiques à Orléans (ESLO) (Dugua et al., 2017 : 48) 70
1.10 Exemples de graphies non standard indiquant la réalisation de liaisons . 78

2.1 Publications scientifiques similaires à l'article *Les Liaisons Dangereuses : Le Statut Équivoque des Erreurs de Liaison* (Desrochers, 1994) 122
2.2 Cartes de l'*Atlas linguistique et ethnographique du Lyonnais* 126
2.3 Exemple de fichier *EUDICO Linguistic Annotator* (ELAN) enrichi (Jean Dumas (JD)_a043_01) 146
2.4 *Où allez-vous jolie-z-et belle que vous marchez si lentement ?* (JD_a060_39) 160
2.5 En-tête du fichier .cha de la chanson *À Vorey le petit bourg* (JD_a361_14) 162

2.6 Vérification du fichier .cha et comparaison avec la fiche manuscrite de Jean Dumas 175

3.1 Virginie Granouillet et Jean Dumas en 1960 182
3.2 Jean Dumas et sa Deux chevaux (2 CV) 189
3.3 Séance d'enregistrement des dentellières par Thelma 189
3.4 Exemples de graphies non standard utilisées par Jean Dumas 191
3.5 Nombre de chansons en top 1 par chanteur 199
3.6 Sexe des chanteurs du corpus de tops 1 199
3.7 Âge des chanteurs du corpus de tops 1 200
3.8 Lieu de naissance des chanteurs du corpus de tops 1 201
3.9 Nombre de tops 1 (partiellement) francophones par année 202
3.10 Nombre de chansons selon la durée de maintien au top 1 202

4.1 Pyramides des âges des différents corpus carottés 205
4.2 Possibilités de requêtes concernant le type de liaison sur l'interface de recherche de Phonologie du français contemporain (PFC) 241
4.3 Enquêtes de PFC où le pataquès *en grand* [*n] *émoi* a été produit dans la tâche de lecture du texte 243
4.4 Enquêtes de PFC où le pataquès *qui mène* [*t] *au village* a été produit dans la tâche de lecture du texte 244

Liste des tableaux

1.1 Paradigmes morphologiques des adjectifs prénominaux 40
1.2 *Murus* : from Latin to Old French (Hornsby, 2020 : 48) 73

2.1 Tableau comparatif des éléments pris en compte par chaque système de codage de la liaison 142
2.2 Variation dans la prononciation des nombres en français contemporain selon leur contexte de réalisation 172

3.1 Corpus compris dans le fonds Jean Dumas 184

4.1 Informations relatives aux corpus carottés 204
4.2 État d'avancement des corpus carottés 206
4.3 Âge des enfants dans le corpus Acquisition de la liaison et interaction parents enfants (ALIPE) à chaque période d'enregistrement (Liégeois, 2014 : 173) 208
4.4 Nombre de pataquès par carotte (modifié depuis Coutanson, 2020 : 9 après réanalyse) 212
4.5 Bilan quantitatif de la première fouille du corpus ALIPE (modifié depuis Coutanson, 2020 : 10 après réanalyse) 239
4.6 Bilan quantitatif de la première fouille du corpus PFC (modifié depuis Coutanson, 2020 : 10 après réanalyse) 246

5.1 Réalisation de liaisons et de pataquès postverbaux dans le fonds Jean Dumas (Coutanson & Badin, sous presse) 276
2 Références des énoncés sélectionnés lors du carottage de l'ALAVAL . . . 300
3 Liaisons codées comme épenthétiques dans le corpus PFC 325

Les exemples oraux de ce manuscrit de thèse sont disponibles à l'adresse suivante : https ://doi.org/10.34847/nkl.bcber4pi.

Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse, Olivier Baude, pour avoir accepté de m'encadrer dans ce projet de recherche, et pour m'avoir permis de candidater à un contrat doctoral dans une université qui n'était encore pas la mienne. Merci Olivier d'avoir toujours été très patient, encourageant et pour le temps très important dégagé pour moi, mais aussi d'avoir toujours cru en ce projet et de l'avoir poussé jusqu'au bout, tout en vérifiant régulièrement que je puisse réaliser cette thèse dans de bonnes conditions.
Je tiens par ailleurs à remercier tout particulièrement les rapporteurs de cette thèse, pour leurs commentaires fournis amenant à réflexion, et pour avoir accepté de lire mon manuscrit dans un délai assez bref. Mes remerciements vont également aux autres examinateurs pour avoir accepté de faire partie de mon jury ainsi que pour la discussion proposée le jour de la soutenance, examinateurs qui pour certains faisaient partie de mon comité de suivi de thèse et qui ont eu l'amabilité de suivre mon parcours au fil des années.

Ce parcours de thèse n'aurait pas été aussi agréable sans tous mes collègues et amis doctorants, postdoctorants et ATER du laboratoire MoDyCo, de notre soutien au quotidien et de l'amitié qui s'est créée entre nous. Dédélienne de cœur devenue modyquienne, je n'oublie pour autant pas mes camarades de promotion lyonnais et leurs sujets de thèse très intéressants, et tous les valeureux jeunes chercheurs rencontrés au fil de mon parcours. Merci aussi aux collègues modyquiens titulaires, notamment à Christophe, mon co-bureau, pour nos discussions très intéressantes et son aide précieuse. J'ai enfin eu la chance de bénéficier de bonnes conditions de travail à MoDyCo, disposant d'un bureau partagé ou encore de la possibilité de partir en mission. Je remercie le personnel administratif de Nanterre, notamment à l'ED 139, au SAS ou les bibliothécaires, pour leur service public quotidien. Je n'oublierai pas les espaces verts, les moutons, dont Zack, ainsi que les murs peints, et l'incroyable vitalité associative de

l’université. Pour finir, j’ai eu la chance de donner cours à des étudiants sympathiques et proposant des sujets d’étude très intéressants.

Enfin, je remercie ma sœur et mon beau-frère qui m’ont permis de m’engager plus sereinement dans ce projet de thèse dans une nouvelle ville, ma sœur à nouveau ainsi que mes parents pour leurs relectures, mes soeurs et mon beau-frère pour s’être admirablement occupés du pot de soutenance, et plus généralement ma famille, mes amis et voisins qui m’ont exprimé leur soutien dans ce projet de longue haleine. Merci à tous ceux qui ont pu être présents le jour de la soutenance ou qui ont pris le temps d’assister à une partie de cette soutenance en ligne, et pour tous les messages de soutien.

Enfin, une pensée pour l’Ukraine, pays dans lequel ce manuscrit a été en partie rédigé, au milieu de quelques alertes aériennes. А ми тую червону калину підіймемо.

À tous ceux qui auraient fait de très belles thèses, et pour qui le destin en a décidé autrement.

Introduction

Cette thèse de doctorat porte sur la liaison, un phénomène morphophonologique qui a fait l'objet de très nombreux travaux en linguistique française, et notamment de nombreuses thèses. Durand et Lyche (2016 : 363-364) en donnent la définition suivante :

> Liaison is a sandhi phenomenon which involves the presence or absence of a consonant at the boundary between two words (Word-1 and Word-2): more specifically, the consonant may appear when Word-2 is vowel initial, but is absent when Word-1 or Word-2 is uttered in isolation.

Pourtant, les linguistes n'ont toujours pas fini de faire le tour de ce sujet d'étude et, loin de s'être tari, celui-ci est continuellement renouvelé par de nouvelles données, de nouvelles analyses et de nouvelles propositions théoriques. Mais un fait linguistique apparenté et marginal demeure pour sa part encore faiblement étudié : il s'agit du phénomène de pataquès, que je définirai p.97 comme des « liaisons inattendues par rapport à la consonne graphique finale du [mot 1] ». J'ai débuté son étude au cours de mes travaux de master, en me penchant sur un objet linguistique n'étant pas étudié à la hauteur de la place qu'il occupe dans notre vie quotidienne : le chant. Ce phénomène nous avait en effet été mentionné lors d'un cours de licence 3, dans le cadre de mes études de Sciences du langage à l'Université Lumière Lyon 2, et je m'étais dit que les pataquès semblaient répandus dans les chansons traditionnelles françaises, où il serait intéressant de les étudier.

Pour mes deux mémoires de master, mon intérêt s'est donc porté tout particulièrement sur le phénomène de pataquès au sein de collections d'exemples concernant la langue chantée. J'avais tout d'abord recensé diverses graphies de pataquès provenant d'une collection d'exemples de paroles de chants traditionnels (Coutanson, 2016 : 70-71, 103-108). J'avais par la suite observé 201 occurrences de pataquès oraux dans une collection d'exemples issus de chansons de tradition orale, qui se répartissaient suivant l'ordonnancement suivant : /z/ > /t/ > /ʁ/ > /n/ = /k/ (Coutanson, 2017 : 73). Je

m'étais aussi intéressée aux pataquès en contexte de liaison, notamment aux consonnes de pataquès remplaçant les consonnes de liaison (Coutanson, 2017 : 76-78), ainsi qu'à quelques contextes spécifiques, comme les pataquès apparaissant entre un auxiliaire et un participe passé ou après un auxiliaire et un participe passé (Coutanson, 2017 : 84-86). J'avais par ailleurs relevé quelques exemples où un pataquès semblait pouvoir être expliqué par l'existence d'une liaison ou d'un pataquès ayant une valeur morphologique dans le même vers d'une autre version de la chanson, ou dans un autre vers de la même chanson (Coutanson, 2017 : 92-93). J'avais en outre ébauché une observation de la répartition spatiale de quelques pataquès issus des vers de mêmes chansons selon le lieu de vie des chanteurs traditionnels (Coutanson, 2017 : 104-110). Enfin, je m'étais intéressée à la réitération des pataquès de chanteurs de l'industrie du disque au cours de leur carrière, et sur la reprise de ceux-ci par d'autres chanteurs reprenant la chanson, dans une collection d'exemples fondée sur les archives de l'Institut national de l'audiovisuel (INA) (Coutanson, 2017 : 112-118).

J'ai décidé de poursuivre mes recherches dans le cadre de cette thèse doctorale en effectuant une étude sur corpus du fonds principal sur lequel je m'étais jusqu'alors penchée, le fonds Jean Dumas (FJD). Celui-ci est composé de chansons de tradition orale recueillies dans le centre de la France entre 1958 et 1962 par un collecteur nommé Jean Dumas et son épouse Thelma. J'étudie ici un échantillon de 498 chansons, d'une durée de 18h 23min, comprenant 89 853 mots pour 5 787 contextes de liaison possibles. Sur cette lancée, je me suis aussi intéressée aux liaisons dans un autre type de corpus chanté, composé de chansons francophones de l'industrie du disque à succès, en compilant un corpus nommé corpus de tops 1, comprenant des titres ayant atteint la première place de classements musicaux réalisés pour la France entre 1956 et 2017. Ce corpus de 368 chansons a une durée de 21h 38min 49s et comptabilise 112 123 mots pour 5 973 contextes de liaison possibles. J'ai par ailleurs eu recours à d'autres types de corpus, où j'ai observé le comportement des phénomènes de liaison et de pataquès : il s'agit du corpus Acquisition de la liaison et interaction parents enfants (ALIPE), un corpus acquisitionnel, interactionnel, dense et recueilli en situation écologique, qui permet notamment d'observer ces phénomènes chez l'enfant ; du corpus Phonologie du français contemporain (PFC), un corpus variationniste enregistré dans divers pays francophones, afin de voir comment se comportent ces phénomènes dans d'autres variétés de français ; de l'Atlas linguistique audiovisuel des dialectes francoprovençaux du Valais romand (ALAVAL), un atlas linguistique semi-spontané de francoprovençal, qui décentre l'étude de ces phénomènes de la langue française et enfin le fonds Charles Joisten, un autre fonds de chansons de tradition orale, lui aussi recueilli dans les années 1960 par un

collecteur nommé Charles Joisten, mais cette fois plus à l'est que le FJD, qui offre un second corpus ethnomusicologique de contrôle.

L'accès à de nouvelles données linguistiques peut à la fois modifier nos certitudes concernant le phénomène de liaison, avec deux corpus chantés dans lesquels on s'attend à rencontrer davantage de liaisons que dans la parole quotidienne, et permettre de dépasser nos connaissances actuelles du phénomène de pataquès, uniquement fondées sur des collections d'exemples bien souvent non attestés. Pour cela, j'ai dû réfléchir à une méthodologie à mettre en place pour analyser ce phénomène rare. Pour ce qui est des corpus PFC et ALIPE, je ne m'intéressais pas vraiment au phénomène de liaison en leur sein, déjà bien étudié, puisqu'ils ont été constitués à cet effet. Par contre, je souhaitais y observer le comportement des pataquès. Je m'attendais à ce que les corpus de chansons de tradition orale ainsi que le corpus acquisitionnel en comportent davantage. Le passage d'une collection d'exemples à une étude sur corpus de ceux-ci m'a notamment permis d'étudier ce phénomène au regard des liaisons présentes dans ce fonds, ce qui n'était pas réalisable à partir de la collection d'exemples recueillis pour mes deux mémoires de master. Dans le cadre d'une linguistique fondée sur l'usage, je me demandais en particulier si les pataquès du corpus pouvaient être expliqués par le fonctionnement des liaisons en leur sein. Pour terminer, le corpus de tops 1 me permettait d'envisager une étude microdiachronique du phénomène de liaison chez de nouveaux professionnels de la parole. Cette thèse constitue donc la première thèse portant sur le phénomène de pataquès, mais ne s'y limite pas. C'est également la première thèse relative au phénomène de liaison dans des corpus chantés.

Dans le premier chapitre de cette thèse de doctorat, je présenterai un état de l'art portant sur ces deux phénomènes morphophonologiques que sont le pataquès et la liaison, d'abord d'un point de vue linguistique, puis sociolinguistique. Dans un deuxième chapitre, la méthodologie que j'ai mise en place sera exposée, en brossant tout d'abord une méthodologie générale, puis en montrant comment cette même méthodologie a été adaptée à l'étude de chaque corpus. Dans un troisième chapitre, le FJD et le corpus de tops 1, qui constituent le gros des nouvelles données pour l'étude de la liaison et du pataquès apportées dans le cadre de cette thèse, feront l'objet d'une présentation détaillée. Les résultats de la deuxième étape de la méthodologie mise en place, soit un carottage de divers corpus sélectionnés et, dès que l'annotation le permettra, une première fouille des données, seront communiqués dans le quatrième chapitre de ce manuscrit. Pour terminer, les résultats de la quatrième étape de cette méthodologie, soit une fouille poussée de mes deux corpus principaux, seront détaillés dans un cinquième et dernier chapitre.

Chapitre 1

Sandhi, liaison et pataquès : d'hyperonyme en hyponyme

1.1 La liaison : un cas de sandhi externe

Le terme ***sandhi*** vient du sanskrit संधि qui signifie « union, association, combinaison ; alliance | jonction, articulation » selon le dictionnaire sanskrit-français *Héritage du sanskrit* (Huet, 2014 : 693). Ce terme était déjà utilisé par les grammairiens de l'Inde ancienne pour décrire des modifications phonétiques caractéristiques du sanskrit. Crystal en propose la définition suivante : "phonological modification of grammatical forms which have been juxtaposed. [...] a distinction is sometimes made between **external sandhi** (sandhi rules which operate across word boundaries) and **internal sandhi** (rules which operate within words)" (Crystal, 1980 : 422). Les sandhis internes étaient nommés [antaraṅga] dans les grammaires de l'Inde ancienne et les sandhis externes [bahiraṅga] (Huet, 2014 : 693). L'anglais est par exemple une langue qui présente à la fois des cas de sandhis internes et externes. On observe un sandhi interne dans l'alternance [k]~[s] entre *electric* et *electricity*. La prononciation diffère dans le deuxième mot seulement car le suffixe *-ity* a été ajouté. Pour présenter un cas de sandhi externe, on peut comparer les groupes nominaux suivants : *a pear* vs *an apple*, dans lesquels le déterminant peut prendre deux formes : *a* et *an*. La première est produite devant initiale consonantique ou glissante tandis que la seconde le sera devant initiale vocalique.

La notion de *sandhi* relève de la morphophonologie[1] puisque la modification phonologique se situe à une frontière morphologique, soit à une frontière de morphèmes

[1] On rencontre les termes *morphophonologie* (c'est celui que j'adopterai ici) ou *morphonologie*. Selon Crystal (1980 : 315), on trouve également le terme *morphophonemics* dans la tradition américaine.

dans le cas des sandhis internes, soit à une frontière de mots dans le cas des sandhis externes. Comme le note Crystal (1980 : 422), un des mérites de ce terme est qu'il peut servir de terme générique recouvrant divers phénomènes morphophonologiques, dont la liaison (voir définition dans le point 1.2.1). Celle-ci, se produisant à la frontière entre deux mots, peut être considérée comme un cas de sandhi externe. Nous verrons néanmoins en 1.2.2.2.2 que le phénomène n'est pas totalement détaché de certains sandhis internes se produisant en français lors de processus de flexion/dérivation. Il est difficile de savoir où borner ce phénomène entre sandhi externe et sandhi interne, cela étant fortement dépendant de notre définition du mot en français (voir aussi Laks, 2005a : 103). Par exemple, les pronoms personnels sujets, malgré leur indépendance graphique, seraient davantage à considérer comme des clitiques selon des critères prosodiques et morphologiques (Blanche-Benveniste, 2010). Dans le cadre de cette étude, les liaisons seront considérées comme telles seulement entre des mots graphiques, excluant de ce fait des toponymes (comme *Vieux-Habitants*), des mots agglutinés (du type *bonhomme* ou *peut-être*), des mots composés (tel *pot-au-feu*) ou encore des sandhis internes (comme *désargenté*), mais incluant des clitiques (tels que *nous* ou *les*) ou encore des locutions figées (comme *de temps en temps*).

Les **sandhis** externes et internes peuvent être **vocaliques** (lorsque la modification est d'ordre vocalique), **consonantiques** (lorsque la modification est d'ordre consonantique : c'est notre cas) ou **tonaux** (lorsque la modification est d'ordre tonale). Ces derniers sont définis par Chen (2000 : Preface : xi) comme des "phonetically conditioned morphotonemic alternations at the juncture of words or morphemes". L'auteur rappelle que la notion a d'abord été appliquée aux langues parlées en Chine et en Asie du Sud-Est avant d'être réutilisée pour des langues parlées sur d'autres continents telles que l'ewe au Ghana, le mixtec au Mexique ou encore le kairi en Papouasie-Nouvelle-Guinée et note que certaines acceptions ont pu inclure des "allotonic variations, intonational effects, and morphologically or syntactically conditioned tone changes" (Chen, 2000 : Preface : xi).

La notion de sandhi est ainsi une notion ancienne utilisée comme hyperonyme pour décrire de multiples phénomènes relevant de la morphophonologie. Les sandhis peuvent être à la fois internes ou externes, et vocaliques, consonantiques ou tonaux. La liaison en français est un cas de sandhi externe consonantique, et je retiens un critère graphique pour délimiter les mots entre lesquels une liaison peut être produite. Tentons de définir maintenant ce qu'est une liaison.

1.2 La liaison : un phénomène complexe

1.2.1 La liaison : définition

Une définition peu sujette à controverse du phénomène de **liaison** car descriptive est donnée par Durand et Lyche (2016 : 363-364) :

> Liaison is a sandhi phenomenon which involves the presence or absence of a consonant at the boundary between two words (Word-1 and Word-2): more specifically, the consonant may appear when Word-2 is vowel initial, but is absent when Word-1 or Word-2 is uttered in isolation.

Par exemple, *deux* se prononcera différemment dans [døzaɲo] pour *deux agneaux* et [døʃəvʁo] pour *deux chevreaux*. On note la présence d'un [z] dans [døzaɲo] qu'on ne retrouve pas dans les formes en isolation : ni dans [dø] pour *deux* ni dans [aɲo] pour *agneaux*. Notons que les liaisons peuvent apparaître également devant glissantes. On en trouve trois en français : /j/ (*des yeux* [dezjø]), /w/ (*des oiseaux* [dezwazo]) et /ɥ/ (*des huîtres* [dezɥitʁ]). Une autre définition intéressante nous est proposée par Bürki et Laganaro (2014 : 169) : « Le terme de liaison décrit le phénomène selon lequel certains mots du français ont deux réalisations (ou variantes), la seconde différant de la première par la présence d'une consonne additionnelle dans certains contextes ». Les deux réalisations dont il est ici question sont deux réalisations phonétiques.

Les études précédentes ont utilisé les formulations suivantes pour se référer aux mots entre lesquels une liaison peut apparaître : ***mot liaisonnant***, ***mot 1*** (M1) ou ***word 1*** (W1) en anglais pour le premier mot, soit le mot précédant la liaison, et ***mot liaisonné***, ***mot suivant***, ***mot déclencheur***, ***mot 2*** (M2) ou ***word 2*** (W2) en anglais pour le second mot, soit le mot suivant la liaison (Chevrot & Fayol, 2001 : 764 ; Durand & Lyche, 2003 : 256, 2008 ; Durand et al., 2011 : 108, 114). La frontière entre ces deux mots constitue ce qu'on nomme un ***contexte de liaison***, un ***site possible de liaison*** ou un ***site potentiel de liaison*** (Durand et al., 2011 : 111, 113).

Pour ce qui est du lien entre liaison et langue française, on peut aisément constater que la plupart du temps, le terme de *liaison* est utilisé pour des études portant sur le français. On peut cependant le trouver dans des études consacrées à des cas de sandhis externes se produisant dans d'autres langues (par exemple pour l'anglais les phénomènes de *linking /r/* et *intrusive /r/* sont regroupés sous l'hyperonyme */r/-liaison* par Mompeán-Gonzalez et Mompeán-Guillamón (2009 : 733)). Il serait néanmoins préférable de réserver le terme *liaison* à des phénomènes liés à celui de liaison en français,

c'est-à-dire portant plus ou moins sur les mêmes consonnes en grande partie étymologiques (ce terme serait donc à utiliser pour des langues romanes, par exemple pour les liaisons qu'on retrouve en francoprovençal). En effet, en généralisant le terme *liaison* à d'autres phénomènes ayant cours dans d'autres langues, celui-ci risque de recouvrir le sens du terme de *sandhi externe*, ce qui est inutile, et nous ne disposerions alors pas d'un terme spécifique pour nommer le phénomène de liaison, ce qui pourrait porter à confusion. Pour les cas en dehors des conditions suscitées, il pourrait ainsi être préférable de se cantonner au terme de *sandhi externe* (consonantique). De la même façon, le terme de liaison en français recouvre plusieurs acceptions. Cela peut porter à confusion lorsque cela concerne d'autres notions en linguistique, comme par exemple la *liaison vocalique* de Grammont (1933 : 417), terme faisant référence à un contact entre deux voyelles sans obstruction.

Nous savons donc qu'une liaison émerge quand deux mots se rencontrent, et que le second est à initiale vocalique ou glissante, mais qu'elle n'apparaît pas si ce dernier est à initiale consonantique ou quand les deux mots sont en isolation. Il serait avantageux de réserver autant que faire se peut le terme de *liaison* à ce phénomène spécifique dans les langues romanes. Faire la liaison n'est pourtant pas systématique. Quand sont produites plus précisément les liaisons ? Il n'existe pas de réponse simple à cette question, car comme l'expliquent Eychenne et Laks (2017 : 1), il s'agit d'un « phénomène multiparamétrique et tous les niveaux linguistiques sont convoqués : phonologie, prosodie et syllabation, morphologie, syntaxe, lexique et sémantique, diachronie, orthographe et différentiation des styles ; [...] sociolinguistique [...] ». Nous nous intéresserons tout d'abord aux facteurs linguistiques influençant la réalisation des liaisons (nous commencerons par le niveau phonologique), avant de nous tourner vers les facteurs sociolinguistiques impliqués dans la réalisation des liaisons.

1.2.2 La liaison : facteurs linguistiques

1.2.2.1 Liaison et phonologie

Comme nous venons de le voir en 1.2.1, la phonologie joue un rôle non négligeable dans le phénomène de liaison puisqu'il est question d'une consonne et de sa présence ou absence dans un contexte spécifique. Prenons le temps de bien distinguer les **consonnes de liaison**[2], dont la réalisation est possible, , uniquement devant initiale vocalique ou

[2]Müller (1985 : 228) utilise le terme de ***liguème*** pour se référer aux phonèmes de liaison.

glissante, mais pas systématique, des **consonnes fixes**, toujours produites, que le mot soit en isolation, qu'il se trouve devant initiale vocalique ou consonantique. Ceci étant dit, nous allons à présent questionner la notion de hiatus souvent convoquée en marge du phénomène de liaison.

1.2.2.1.1 Liaison et hiatus

Le **hiatus** « correspond à la rencontre de deux voyelles appartenant à des syllabes différentes » (Mallet, 2008 : 48). Faisons un petit aparté sur un débat récurrent : faut-il faire la liaison et l'élision avec le mot *hiatus* ? On dispose aujourd'hui des moyens nécessaires pour étudier l'évolution diachronique de l'élision de ce terme à l'écrit. Dans la figure 1.1[3], on remarque qu'en dépit de ce que peuvent affirmer les puristes, il y a toujours eu de la variation. Il est vrai que les formes élidées étaient dominantes jusqu'il y a peu mais elles semblent aujourd'hui en perte de vitesse face aux formes non élidées. Notons par ailleurs que la tendance s'est inversée un peu plus tôt pour le couple *d'hiatus/de hiatus* que pour le couple *l'hiatus/le hiatus*.

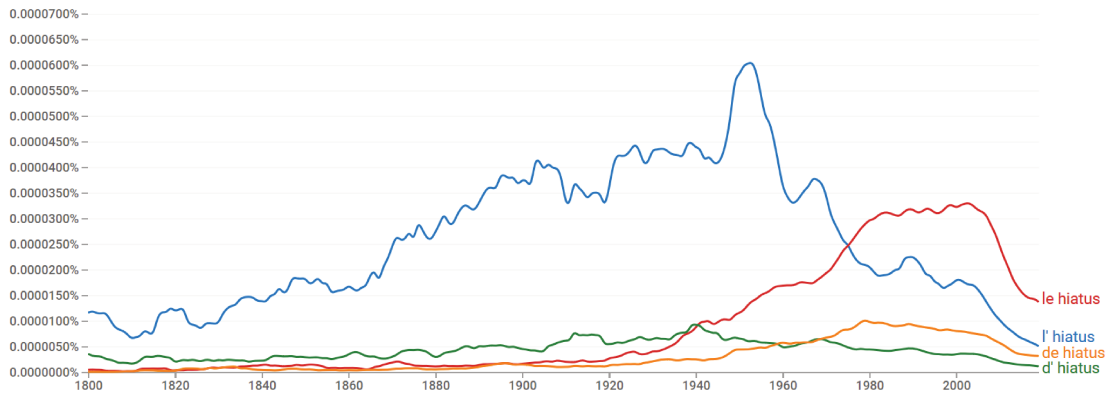


FIGURE 1.1 : Élision du mot *hiatus* dans le corpus French 2019 (Google Ngram Viewer)

Pendant un temps, il a été dit que la production d'une liaison constituait une stratégie anti-hiatus mise en place pour éviter les séquences VV (voyelle + voyelle), peu appréciées en français. Dans notre cas, il est question d'une finale vocalique rencontrant

[3]Graphique obtenu le 30 mars 2021 en utilisant le Google Books Ngram Viewer (Michel et al., 2011), accessible en ligne à l'adresse suivante : https://books.google.com/ngrams. Le corpus French 2019 (1800-2019) de livres Google a été sélectionné, sans avoir opté pour une insensibilité à la casse et avec smoothing = 3 (valeur par défaut, ne déplaçant pas le point temporel d'inversion des tendances comparé à smoothing = 0 tout en proposant une meilleure lisibilité). Notons que le terme *hiatus* ne réfère pas toujours dans les livres Google au phénomène linguistique que nous abordons ici.

une initiale vocalique lors de la succession de deux mots : les hiatus qui poseraient problème sont ainsi les **hiatus** dits **transitoires**, ceux se produisant à la frontière entre des mots (exemple 1). Il existe aussi des **hiatus** dits **internes**, se trouvant entre deux syllabes (comme avec [aunuprodɥi] pour *août nous produit* dans l'exemple 83 ; cf Mallet, 2008 : 48).

(1) *SGR : mon tendre coeur galant n'est pas pour vous je l'ai promis à un autr(e) avant vous.
%pho : mɔ̃ tɑ̃dʁə kœʁ galɑ̃ nɛ pa puʁ vuː ʒə lɛ pʁomi a œ̃ notʁ avɑ̃ vuː
(JD_a347_06, 43200, Annette Rousset)

Morin (2005 : 8) nous rappelle que l'évocation de cette fonction anti-hiatique provient du discours esthétique sur la poésie. Mais cet argument est par exemple repris plus récemment dans le cadre de la Théorie de l'Optimalité (OT), avec la mention d'une contrainte anti-hiatus notée *VV[4] (Tranel, 2000 : 53). Dans son article *La liaison relève-t-elle d'une tendance à éviter les hiatus ? Réflexions sur son évolution historique*, Morin (2005) tente de démonter ce mythe persistant et maintes fois repris. Voici les arguments que nous retiendrons :

- la non-prise en compte de tous les contextes de liaison (Morin, 2005 : 16). Comme le note plus explicitement Mallet (2008 : 54), l'argument de la fonction anti-hiatique de la liaison semble découler de l'étude prioritaire des contextes de liaisons invariables qui a pu éclipser les cas de liaisons variables non réalisées qui allaient pourtant à l'encontre de cet argument.
- la non-réalisation de la liaison après les adjectifs masculins singuliers prénominaux en wallon de Liège, alors qu'elle est réalisée au pluriel (Morin, 2005 : 16-17). Dans ce dialecte du wallon (une langue d'oïl), les adjectifs sont habituellement antéposés au nom. Si la liaison découle d'une stratégie anti-hiatus, elle devrait s'appliquer autant après des adjectifs singuliers que pluriels. Le dialecte liégeois du wallon a été invoqué afin de parer à la spécificité du français, où un nombre restreint d'adjectifs précède un nom à l'oral : seuls 31 lemmes pouvant liaisonner

[4] *VV signifie que les séquences voyelle + voyelle sont bannies dans les langues du monde (dans la théorie de l'optimalité, les contraintes sont universelles mais sont hiérarchisées différemment selon les langues, les dialectes ou encore les registres). Si un **candidat** (c'est-à-dire un output potentiel, ou réalisation phonétique potentielle) présente une séquence voyelle + voyelle, on dit que la contrainte est transgressée. Dans son article *Aspects de la phonologie du français et la théorie de l'optimalité*, Tranel (2000 : 58) présente la contrainte *VV comme dominée, ce qui signifie que si une contrainte dominante est transgressée par un candidat, celui-ci sera automatiquement éliminé. De cette façon, dans ce cadre théorique, la contrainte anti-hiatus interfère dans la réalisation des liaisons, mais de façon moindre.

ont été observés dans ce contexte par Barreca et al. (2015) dans de l'oral spontané (corpus PFC et Corpus de français parlé parisien des années 2000 (CFPP2000)). En français, ces adjectifs antéposés étant d'ordinaire liaisonnants (sans différence significative de réalisation qu'ils soient singuliers ou pluriels dans les corpus étudiés par Barreca et al., 2015 : respectivement liaisonnés à 84% et 86%), cet argument est inopérant. Reste à savoir si une étude sur corpus du dialecte liégeois apporterait les mêmes résultats.

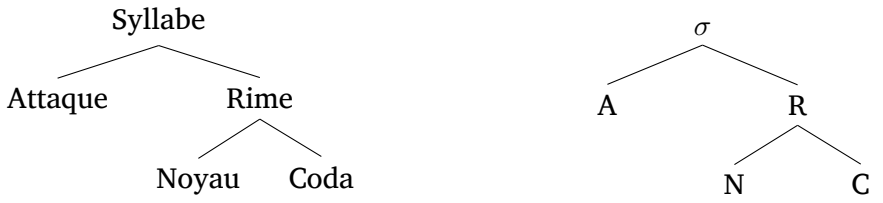
- la réalisation de liaisons entre une finale consonantique et une initiale vocalique (Morin, 2005 : 19). Si une contrainte anti-hiatus était réellement à l'œuvre, on ne devrait pas observer de liaisons dans un tel contexte, puisque les liaisons ne devraient apparaître que pour éviter un hiatus entre finale vocalique et initiale vocalique.
- la non-adéquation avec nos connaissances en matière d'évolution du français, où des consonnes ne sont pas apparues pour éviter qu'il y ait hiatus, mais au contraire, ont disparu devant une autre consonne (Morin, 2005 : 18). On pourrait alors penser qu'elles se sont maintenues pour empêcher des rencontres inopportunes de voyelles. Cela rejoint le premier argument cité, car dans ce cas-là, les liaisons auraient été conservées dans toutes les positions où il y avait un risque de hiatus, ce qui n'est pas le cas. Morin, citant Pope (1952) et Bybee (2000), nous explique que la variante préconsonantique, où la consonne a chuté, se serait étendue aux contextes prévocaliques en raison de fortes pressions cognitives sur l'unicité des représentations (Morin, 2005 : 18-19).

Le seul argument qui pourrait aller dans le sens d'une stratégie anti-hiatus pour la liaison serait le cas des liaisons non étymologiques, toutefois leur apparition erratique n'en fait pas un argument de poids. Si la force anti-hiatique de la liaison était telle, elles auraient dû être introduites à chaque rencontre de voyelle à la frontière entre deux mots. En outre, cette raison phonologique n'apporte aucune explication quant à la nature de la consonne de liaison devant être insérée entre les deux voyelles.

On rencontre ainsi des hiatus transitoires ou internes. L'argument fréquemment avancé selon lequel faire la liaison permettrait d'éviter les hiatus transitoires, trouvant son origine dans la poésie et remis au goût du jour dans les travaux (post-)génératives, va à l'encontre de l'évolution diachronique de la liaison, de données en diatopie à son égard ou est encore remis en cause par la non-unicité de ce phénomène. Cet argument ne semble donc pas pertinent dans l'étude de la liaison. Intéressons-nous maintenant à la syllabation de la consonne de liaison.

1.2.2.1.2 Liaison et syllabation

Le phénomène de liaison doit bien être distingué du phénomène d'**enchaînement**. Cette distinction est proposée au moins depuis Pernot (1937). Le schéma de la structure syllabique en 1.2 éclairera ce qui suivra.



(a) Présentation développée (b) Présentation abrégée

FIGURE 1.2 : Schéma de la structure syllabique

L'enchaînement correspond à la resyllabation de la consonne finale d'un M1 (coda de la dernière syllabe du M1) à l'initiale d'un M2 (attaque de la première syllabe du M2) : on parle de resyllabation car la consonne en question n'est plus prononcée au sein de la dernière syllabe du M1, comme dans la forme produite en isolation, mais appartient à la première syllabe du M2. Les exemples suivants sont issus de la tâche de lecture du corpus PFC pour une même locutrice. Il existe :

- des non-enchaînements ne correspondant pas à des cas de liaison, impliquant des consonnes fixes n'étant pas resyllabées devant initiale vocalique (dans l'exemple 2, il s'agit du [t] final de *côte* ou du [n] final de *mène*).

  (2) La côte escarpée du Mont Saint-Pierre qui mène au village [...].
  [la.kot.ʔɛs.kaʁ.pe.dy.mɔ̃.sɛ̃.pjɛʁ.ki.mɛn.ʔo.vi.laʒ]
  (PFC, Nyon (Suisse, Europe), 2002, svarb2tw)

- des enchaînements ne correspondant pas à des cas de liaison, impliquant des consonnes fixes resyllabées devant initiale vocalique (cela peut être illustré par la prononciation du [l] final de *il* devant *a* ou du [n] final de *une* devant *impasse* dans l'exemple 3). L'enchaînement en français est habituellement réalisé.

  (3) Il a le sentiment de se trouver dans une impasse stupide.
  [i.la.lə.sɑ̃.ti.mɑ̃.də.sə.tʁu.ve.dɑ̃.zy.nɛ̃.pɑs.sty.pid]
  (PFC, Nyon (Suisse, Europe), 2002, svarb2tw)

- des cas de **liaisons enchaînées** (ou **avec enchaînement**), impliquant des consonnes de liaison resyllabées devant initiale vocalique (voir par exemple la pronon-

ciation de *est* devant *en* ou de *très* devant *inquiet* dans l'exemple 4). Ces cas constituent la majorité des liaisons relevées en corpus : ces liaisons sont prototypiques.

(4) Le maire de Beaulieu - Marc Blanc - est en revanche très inquiet.
[lə.mɛʁ.də.bo.ljø.maʁk.blɑ̃.ɛ.tɑ̃.ʁə.vɑ̃ʃ.tʁɛ.zɛ̃.kjɛ]
(PFC, Nyon (Suisse, Europe), 2002, svarb2tw)

- des cas de **liaisons non enchaînées** (ou **sans enchaînement**), notamment étudiées par Encrevé (1983, 1988)[5], impliquant des consonnes de liaison n'étant pas resyllabées devant initiale vocalique : une consonne de liaison est prononcée comme la consonne finale du M1, puis on trouve principalement une pause et/ou une occlusive glottale, plus rarement un schwa ou la marque d'hésitation *euh*, avant la voyelle initiale du M2 (voir la prononciation de *s'est* devant *en* dans l'exemple 5).

(5) Il s'est, en désespoir de cause [...].
[il.sɛt.ʔɑ̃.dɛ.zɛs.pwaʁ.də.koz]
(PFC, Nyon (Suisse, Europe), 2002, svarb2tw)

Si ces liaisons non enchaînées ont été notées de façon significative par Encrevé chez les personnalités politiques, il n'en est pas de même pour les corpus de locuteurs non professionnels, notamment PFC (Durand et al., 2011 : 114). Notons que chez les sujets d'étude d'Encrevé, c'étaient en écrasante majorité les contextes variables qui étaient concernés : sur 5027 liaisons invariables réalisées, seules 11 n'ont pas été enchaînées, soit 0,2%, alors que sur 2815 liaisons variables réalisées, 316 n'ont pas été enchaînées, soit 11,2% (Encrevé, 1983 : 51). La variation inter-individuelle était en outre importante chez les personnalités politiques qu'il a étudiées (de 0% de liaisons non enchaînées pour M. Rocard à 18,7% pour M. Debré ; Encrevé, 1983 : 49), tout comme la variation intra-individuelle, que ce soit entre les diverses productions d'un même locuteur (par exemple, V. Giscard d'Estaing produisait entre 8,2% et 33,3% de liaisons sans enchaînement dans les discours qu'a étudié Encrevé ; Encrevé, 1983 : 51) ou à l'intérieur d'une même production (Encrevé note que les liaisons non enchaînées « apparaissent le plus souvent en série » ; 1983 : 60).

[5] À la même époque, Milner et Regnault, dans leur ouvrage *Dire le vers* (1987 : 55-56), mentionnent eux aussi ces deux prononciations pour la liaison. Ils parlent de ***diction directe*** pour les liaisons enchaînées et de ***diction indirecte*** pour les liaisons non enchaînées. Ils notent ces dernières en exposant du M1. Pour ces auteurs, la diction indirecte est une « diction allégée de la liaison » (Milner & Regnault, 1987 : 66). Laks (2008 : 237-238) présente encore d'autres auteurs mentionnant le phénomène de manière anecdotique.

En résumé, il ne faut pas confondre liaison et enchaînement. Une consonne fixe peut être enchaînée ou non, tout comme les consonnes de liaison. Dans l'exemple 4, le [t] de liaison de *est* est resyllabé avec le mot suivant, *en* : il s'agit d'une liaison enchaînée. Au contraire, dans l'exemple 5, le [t] de liaison de *s'est* n'est pas resyllabé au mot suivant, *en* : il s'agit d'une liaison non enchaînée. Cette dernière a surtout été notée chez les hommes politiques, avec beaucoup de variations inter- et intra-individuelles et concerne quasi-exclusivement les liaisons variables. Quelles consonnes retrouve-t-on en liaison ? C'est ce sur quoi nous nous pencherons dans la sous-partie suivante.

1.2.2.1.3 Nature et prégnance des consonnes de liaison

1.2.2.1.3.1 Nature des consonnes de liaison au niveau phonologique

Au niveau phonologique, on trouve les consonnes de liaison suivantes en français : /p, t, n, z, ʁ/ (Figure 1.3). La liaison va donc impliquer des occlusives non voisées : actuellement /p/ et /t/, des fricatives voisées : /z/ et //ʁ//[6] et la nasale /n/. Comme le remarquent Durand, Laks et al. (2011 : 125), ces consonnes ne constituent pas une classe naturelle. Quatre autres consonnes peuvent être mentionnées dans la littérature : /k, g, v, l/, que je ne considère pas relever du phénomène de liaison, même s'il est bien question ici de sandhis consonantiques : d'autres sandhis externes du français impliqueront la fricative voisée /v/ ou encore l'approximante latérale /l/, et des sandhis internes impliqueront enfin, entre autres, l'occlusive non voisée /k/.

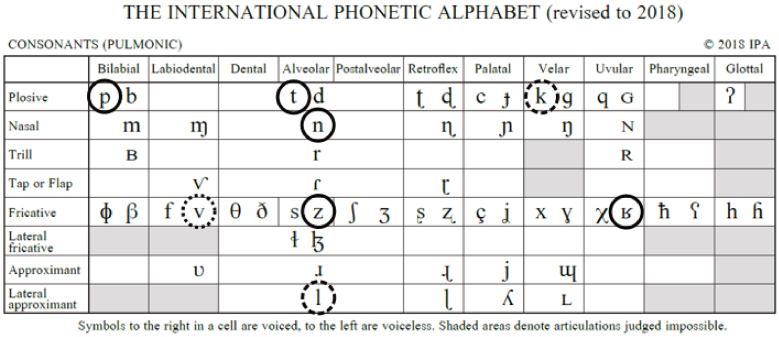


FIGURE 1.3 : Consonnes de liaison du français et autres consonnes impliquées dans des sandhis consonantiques

Pour ce qui est de la consonne /k/, on ne trouve de nos jours des attestations d'un

[6] La notation entre double barres obliques indique qu'il s'agit d'un diaphonème (voir 1.2.2.1.3.2).

‹g› prononcé [k] que dans des toponymes comme *Bourg-en-Bresse* [buʁkɑ̃bʁɛs] ou *Bourg-Argental* [buʁkaʁʒɑ̃tal]. Cela pourrait à première vue s'apparenter à des cas de liaison puisque dans d'autres toponymes, devant consonne, ce ‹g› n'est pas prononcé comme dans Bourg-la-Reine [buʁlaʁɛn]. Néanmoins, ceci serait plutôt à considérer comme des cas de sandhis internes, ce que corrobore la présence de semblables réalisations dans des mots composés tels que *blanc-étoc/blanc-estoc* [blɑ̃ketɔk]/[blɑ̃kɛstɔk], *croc-en-jambe* [kʁokɑ̃ʒɑ̃b], *franc-alleu* [fʁɑ̃kalø] ou *porc-épic* [pɔʁkepik] (Ahmad, 1993 : 40-41) ou encore dans des locutions adverbiales comme *à franc étrier* [afʁɑ̃ketʁije] (CNRTL, 2012c), alors que les mots *blanc*, *croc*, *franc* et *porc* ne présentent pas de consonne finale en isolation. La liaison étant apparentée à un cas de sandhi externe, /k/ ne peut être considéré à l'heure actuelle comme une consonne de liaison. Dans le corpus PFC, aucune liaison en /k/ n'a été observée (Durand et al., 2011 : 124). On a pu lire dans la littérature que parfois, une liaison en /k/ était produite de façon voisée, soit [g] (cf par exemple Carton, 1974 : 88 ; Morin, 1987 : 819). De telles occurrences n'ont de même pas été observées dans PFC. Dispose-t-on réellement de telles attestations hors tâche de lecture ou écrits oralisés ?

Pour ce qui est de la consonne /l/, elle pourrait également être considérée à première vue comme une consonne de liaison. En effet, les pronoms personnels sujets de troisième personne du singulier *il* et *elle* peuvent prendre deux formes phonétiques en français : [il] et [ɛl] mais également [i] et [ɛ] devant certaines initiales consonantiques (vers (6) et (9))[7]. Si [il] et [ɛl] sont toujours produits devant initiale vocalique en français hexagonal (vers (7) et (10)), on ne peut pas réellement parler d'un cas de liaison car les formes à consonnes finales se retrouvent souvent devant initiales consonantiques (vers (8) et (11)). C'est ici le manque de systématicité du phénomène qui ne m'amène pas à considérer cette consonne comme de liaison, même s'il s'agit bien d'un sandhi externe consonantique.

(6) *SGR : s'il en rest(e) pas du tout e(lle) s'en bross(e)ra le ventre tambour battait.
(JD_a159_02, 43800, André-Joseph Aguilhon)

(7) *SGR : ell(e) en boira son plein saoul la gueritaine.
*SGR : ell(e) en boira son plein saoul et voilà tout.
(JD_a347_03, 43200, Séraphine Chaize)

[7] Selon Morin (1982 : 31), le schwa final de *elle* aurait empêché dans un premier temps l'amuïssement du [l] du pronom, contrairement à celui de *il* qui se serait amuï plus précocement. Cela pourrait expliquer pourquoi on rencontre davantage la forme préconsonantique courte pour *il* que pour *elle* en français hexagonal.

(8) *SGR : à chaqu(e) instant ell(e) me cherche querell(e).
*SGR : les pots les verr(e)s tout se casse sur moi.
(JD_a060_08, 43810, Virginie Granouillet)

(9) *SGR : le plu(s) petit dit à les autr(e)s i(l) faut la prendre l'emporter.
(JD_a060_34, 43810, Virginie Granouillet)

(10) *SGR : mes amis vous pouvez le croir(e) il est pir(e) qu'un.
*SGR : lion il est pir(e) qu'un lion.
(JD_a062_06, 43810, Virginie Granouillet)

(11) *SGR : de suit(e) il s'est mis à lui dir(e) de suit(e) il s'est mis à lui dir(e).
(JD_a338_03, 43200, Séraphine Chaize)

Enfin, on lit parfois dans la littérature que le /v/ serait une consonne de liaison (Carton, 1974 : 88 ; Malécot, 1975 : 168 ; Müller, 1985 : 228). Il est vrai que dans de très rares cas un ‹f› final, prononcé généralement [f], se voise devant initiale vocalique, comme pour les principaux exemples cités *neuf ans* [nœvɑ̃] et *neuf heures* [nœvœʁ] (Delattre, 1947 : 151). Si nous sommes bien face à un cas de sandhi externe, ce n'est pas de la liaison car il n'y a pas (plus[8]) d'alternance entre présence et absence d'une consonne.

Seront dès lors considérées comme consonnes de liaison les consonnes suivantes en français standard contemporain : /p, t, n, z, ʁ/. /k/ sera à réserver pour décrire un sandhi interne du français (visible dans des toponymes, mots composés, locutions adverbiales) et /v, l/, n'étant pas assez systématiques, ne correspondent qu'à d'autres sandhis externes de cette langue. /g/ me semble discutable hors tâche de lecture. Les consonnes retenues correspondent à des phonèmes du français. Nous nous demanderons dès lors si ces consonnes admettent différentes réalisations phonétiques.

1.2.2.1.3.2 Nature des consonnes de liaison au niveau phonétique

Au niveau phonétique, nous n'avons pas l'habitude de rencontrer beaucoup de variation allophonique en français en ce qui concerne les liaisons. Tout au plus, on peut observer de la variation contextuelle, comme dans les exemples suivants issus pour le premier du corpus PFC et pour le second du FJD, où une harmonie consonantique a été produite par persévération d'une consonne :

(12) Chaque fois il y a des [z] étrangers, même chez nous, chaque fois j'ai des [ʒ] ét/, des [z] étrangers.

[8] voir exemples 105, 106, 107 et 108

(PFC, Bamako (Mali, Afrique subsaharienne), 2006, maazw1)

(13) *SGR : bonjour la compagnie sans [ʒ] oublier ma mie.
(JD_a165_09, 43800, Germaine Varenne)

On peut aussi noter différentes réalisations du phonème /ʁ/ en français telles que [ʁ], [ʀ] ou encore [r]. C'est pour cela que j'opte pour la notation entre double barres obliques (//ʁ//), afin d'indiquer qu'il s'agit d'un diaphonème.

Ce n'est pas le cas en francoprovençal, une langue régionale parlée entre la France, la Suisse et l'Italie. Dans l'*ALAVAL*, on note par exemple une grande variation allophonique correspondant au même diaphonème :

(14) [do we k lu ð ɑ̃fɑ̃ː - aʎɑ̃ døʁmˈi ɐvˈɑ̃ nʊ ˈʊʀə]
Je veux que les enfants .. aillent dormir avant neuf heures.
SixtM : Je veux que les enfants se couchent avant neuf heures.

(15) [lø ʐ̩[9] avˈœʎə ɪ faŋ də bʁˈeiːdə pə l ɑʊəʎˈiː]
Les abeilles elles font de bruit dans la ruche.
BionazM : Les abeilles bourdonnent dans la ruche.

(16) [kã laʁə gamˈinə lɑ̃mˈɔvə pə ɪ z epinˈa]
Quand j'étais gamine j'aimais pas les épinards.
ChamosonF : Quand j'étais gamin/gosse, je n'aimais pas les épinards.

(17) [ɪ kɔ̃tˈɑː na dɔzˈɑ̃ŋnɑ ə də ʒ ɐrɑ̃ndˈɑlː ɐ̟nˈa ʃʊ lʏ fek]
J'ai compté une douzaine euh de-z-hirondelles en haut sur les fils.
HérémenceM : J'ai compté une douzaine d'hirondelles sur le fil.

(18) [lɐ̟ e ʑ ɥøs pˈiːrə]
Elle a les yeux bleus.
ArbazF : Elle a les yeux bleus.

La variation allophonique touchant les consonnes de liaison semble se limiter en français à de rares cas d'harmonies consonantiques et aux différentes réalisations du diaphonème //ʁ//, contrairement au francoprovençal pour lequel j'ai exemplifié de multiples allophones pour l'équivalent de la liaison en /z/ du français. Dans le point suivant, nous allons aborder la question de la fréquence d'apparition des divers segments de liaison, en recensant les multiples ordonnancements qu'on rencontre dans la littérature.

[9]La transcription phonétique de l'ALAVAL est faite en Alphabet phonétique international (API) adapté. Le symbole [ʐ̩] correspond en API à [ʒ̩] (Berchtold et al., 2019b : 26).

1.2.2.1.3.3 Prégnance des consonnes de liaison

Revenons au cas du français. Toutes les consonnes de liaison ne sont pas aussi prégnantes, notamment, les liaisons en //ʁ// et /p/ sont extrêmement rares, quels que soient les corpus (voir par exemple ce qu'il en est pour le corpus PFC dans le tableau de la figure 1.4). Il faut dire que les contextes après lesquels peuvent apparaître ces liaisons sont restreints (et sont tous des contextes variables) : ceux-ci sont présentés en 1.2.2.4. Il en irait autrement pour la consonne /n/, puisqu'il a été noté dans la littérature que ces liaisons appartiendraient principalement à des contextes invariables (Durand et al., 2011 : 125; Adda-Decker et al., 2012 : 122). Quant aux consonnes de liaisons /z/ et /t/, elles appartiennent à la fois à des contextes invariables et à des contextes variables.

Pour ce qui est des liaisons les plus fréquemment observées en corpus (voir figure 1.4[10]), deux ordonnancements sont mentionnés dans la littérature :

- /z/ > /t/ > /n/ > /p/ > //ʁ// pour les corpus de Malécot (1975 : 168)[11] et d'Ahmad (1993 : 98)[12]. Notons que le corpus d'Ahmad comprend trois sous-corpus, dont le corpus de Malécot qui en constitue la part la plus importante (4909 liaisons réalisées/6312; Ahmad, 1993 : 96).
- /z/ > /n/ > /t/ > //ʁ// > /p/ pour le corpus PFC (Durand et al., 2011 : 124; cf Durand & Lyche, 2008 pour une description du corpus), de même /z/ > /n/ > /t/ pour le NCCFr (Adda-Decker et al., 2012; cf Torreira et al., 2010 pour une description du corpus) et /z/ > /n/ > /t/ > //ʁ// pour le module Diachronie du corpus des ESLO (Dugua et al., 2017 : 50; cf Eshkol-Taravella et al., 2011 pour une description du corpus).

Dugua et Baude (2017 : 51) observent pour le module Diachronie du corpus des ESLO que l'ordre de fréquence des liaisons invariables est le suivant : /z/ > /n/ > /t/ et que celui des liaisons variables est le suivant : /t/ > /z/ > /n/ > /ʁ/. Cette remarque est très

[10] Seuls les effectifs du tableau de la figure 1.4 étaient fournis dans les articles pour le corpus de Malécot, PFC, le *Nijmegen Corpus of Casual Speech* (NCCFr) et les ESLO, seuls les pourcentages étaient fournis pour le corpus d'Encrevé.

[11] Ont ici été retirés des résultats de Malécot les /v/ (3), les /l/ épenthétiques (4) ainsi que le /t/ épenthétique (1).
L'article de Malécot n'est pas explicite sur le fait de savoir si ce sont les liaisons réalisées ou les contextes de liaisons qui sont présentés pour chaque consonne. Notamment, il mentionne "4,409 possible liaisons", ce qui correspond à la somme des divers effectifs du tableau de la figure 1.4 et qui ferait plutôt pencher du côté de contextes de liaison. Pourtant, lorsqu'Ahmad (1993 : 98) dénombre les liaisons réalisées pour son sous-corpus M (Malécot), les chiffres sont encore supérieurs, ce qui laisse à penser qu'il s'agissait bien de liaisons réalisées et que tout le corpus n'avait pas été traité dans l'étude de Malécot (1975).

[12] De même, pour Ahmad (1993 : 94,98) le nombre de liaisons réalisées manque de clarté : plusieurs chiffres par consonne de liaison et pour le total sont avancés.

pertinente, car les différents ordonnancements notés dans la littérature pourraient dépendre du taux de réalisation des liaisons variables du corpus. Enfin, en ce qui concerne exclusivement les liaisons non enchaînées, mentionnées en 1.2.2.1.2, Encrevé (1983 : 51) a montré que l'ordonnancement était le suivant dans son corpus de personnalités politiques : /t/ > /z/ > /ʁ/ > /n/ > /p/. Une des explications pourrait être que les liaisons non enchaînées concernent presque exclusivement des liaisons variables (voir 1.2.2.1.2).

Suivant la réflexion engagée par Mallet (2008), il semble nécessaire de bien savoir ce qui est recherché derrière chaque ordonnancement proposé :

- la part de réalisation potentielle de chaque consonne de liaison dans l'ensemble des contextes de liaison (figure 1.5) ? Dans ce cas, pour PFC l'ordonnancement des contextes de liaison est le suivant : /z/ > /t/ > /n/ > /ʁ/ > /p/ > /k/. Il en va de même pour le corpus NCCFr : /z/ > /t/ > /n/[13]. Ce serait l'ordonnancement des consonnes de liaison en français si celles-ci étaient toutes prononcées.
- la part de réalisation de chaque consonne de liaison dans l'ensemble des liaisons réalisées (figure 1.4) ? C'est généralement ce qu'on entend par ordonnancement des consonnes de liaison et ce que je viens de présenter au début de cette section.
- la part de réalisation de chaque consonne de liaison dans l'ensemble des réalisations potentielles de cette même liaison ? Dans ce cas, pour le corpus de Malécot (1975 : 168), PFC (Mallet, 2008 : 213) et le NCCFr (Adda-Decker et al., 2012 : 122), l'ordonnancement observé ferait consensus pour les trois consonnes de liaison les plus fréquentes : /n/ > /z/ > /t/ (tableau 1.6).

Pour conclure, si on s'intéresse aux liaisons réalisées, deux ordonnancements des consonnes ont été observés selon les corpus : /z/ > /t/ > /n/ ou /z/ > /n/ > /t/, différence qui semble due au poids des liaisons variables réalisées pour chaque corpus. En raison de contextes d'apparition restreints, les liaisons en //ʁ// et /p/ restent erratiques. Pour les liaisons non enchaînées, /t/ > /z/ semble de rigueur. Si on se penche à présent sur la part de réalisation potentielle de chaque consonne de liaison, l'ordonnancement par défaut semble être /z/ > /t/ > /n/. Pour finir, si on regarde de plus près la productivité de chaque consonne de liaison, l'ordonnancement /n/ > /z/ > /t/ semble faire l'unanimité, ce qui s'explique par le fait que la plupart des liaisons en /n/ sont

[13] Les chiffres présentés ici ne reflètent pas tout à fait la réalité car dans l'article concernant le NCCFr, seule une approximation du nombre de contextes de liaison pour /ʁ/ et /p/ est fournie, ce pourquoi les données concernant ces consonnes ont ici été omises.
Seuls les effectifs du tableau de la figure 1.5 étaient fournis dans les articles pour PFC et le NCCFr.

Dugua et al., 2017 : 50	z (606 ; 49,31%)	n (429 ; 34,91%)	t (192 ; 15,62%)	ʁ (2 ; 0,16%)	p (0 ; 0%)
Malécot, 1975 : 168	z (2447 ; 55,60%)	t (1173 ; 26,65%)	n (769 ; 17,47%)	p (8 ; 0,18%)	ʁ (4 ; 0,09%)
Ahmad, 1993 : 98	z (3195 ; 50,64%)	t (1918 ; 30,40%)	n (1180 ; 18,70%)	p (11 ; 0,17%)	ʁ (5 ; 0,08%)
PFC (Durand et al., 2011 : 124)	z (11000 ; 45,92%)	n (8515 ; 35,55%)	t (4133 ; 17,25%)	ʁ (42 ; 0,18%)	p (14 ; 0,06%)
PFC, Bangui (Bordal, 2012 : 37)	z (490 ; 58,0%)	t (222 ; 26,2%)	n (134 ; 15,8%)	NA	NA
PFC, Neuchâtel (Racine & Andreassen, 2012 : 199)	n (223 ; 43,73%)	z (209 + 12 ; 43,33%)	t (66 ; 12,94%)	NA	NA
NCCFr (Adda-Decker et al., 2012 : 122)	z (3958 ; 48,31%)	n (3247 ; 39,63%)	t (7530 ; 12,06%)	NA	NA
Encrevé, 1983 : 51	t (77,5%)	z (17,5%)	ʁ (3%)	n (1,5%)	p (0,5%)
FJD, ce (Coutanson, 2017 : 73)	z (110 ; 54,73%)	t (67 ; 33,33%)	ʁ (20 ; 9,95%)	n (3 ; 1,49%)	k (1 ; 0,5%)

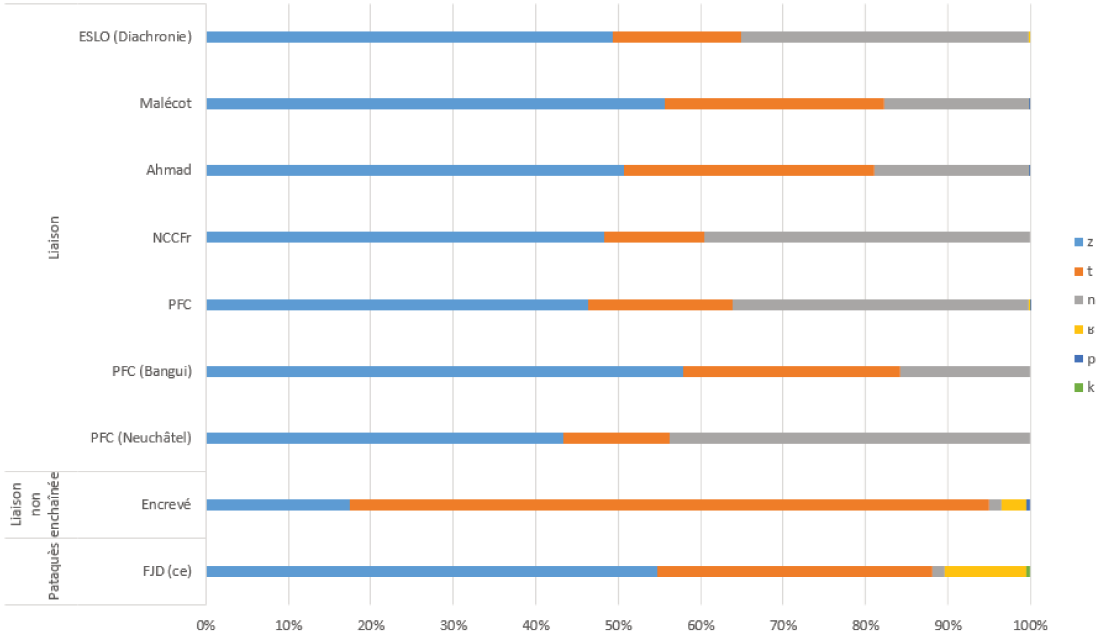


FIGURE 1.4 : Part de réalisation de chaque consonne de liaison ou de pataquès dans l'ensemble des liaisons ou pataquès réalisés

PFC (Mallet, 2008 : 213)	z (13852 ; 48,57%)	t (8113 ; 28,45%)	n (5512 ; 19,33%)	ʁ (942 ; 3,30%)	p (98 ; 0,34%)	k (2 ; 0,01%)
NCCFr (Adda-Decker et al., 2012 : 122)	z (14589 ; 54,14%)	t (7530 ; 27,94%)	n (4827 ; 17,91%)	NA	NA	NA

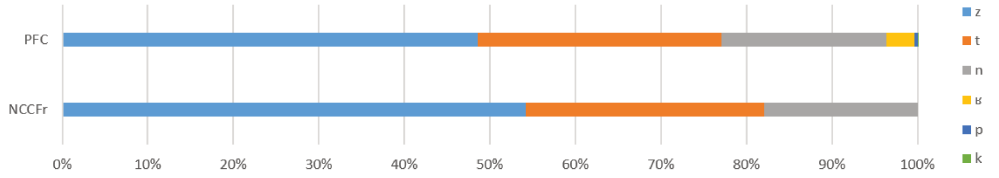


FIGURE 1.5 : Part de réalisation potentielle de chaque consonne de liaison dans l'ensemble des contextes de liaison

Malécot, 1975 : 168	n (94,4% de n = 769)	z (60,5% de n = 2447)	t (52,4% de n = 1173)	ʁ (50% de n = 4)	p (25% de n = 8)
PFC (Mallet, 2008 : 213)	n (4970/5512 ; 90,2%)	z (5960/13 852 ; 43%)	t (1888/8113 ; 23,3%)	p (11/98 ; 11,2%)	ʁ (12/942 ; 1,3%)
NCCFr (Adda-Decker et al., 2012 : 122)	n (3247/4827 ; 67%)	z (3958/14589 ; 27%)	t (988/7530 ; 13%)	NA	NA

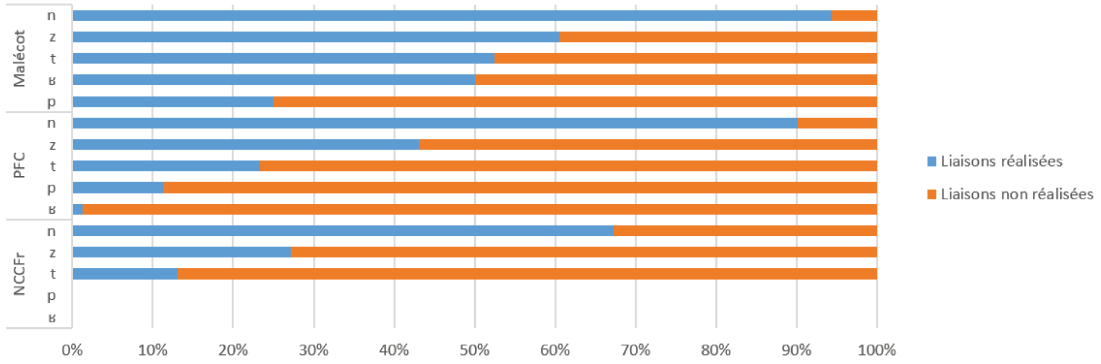


FIGURE 1.6 : Part de réalisation de chaque consonne de liaison dans l'ensemble des réalisations potentielles de cette même liaison

invariables. Maintenant que nous avons une bonne vue d'ensemble du phénomène de liaison sous l'angle phonologique, il ne nous reste plus qu'à nous arrêter sur les segments adjacents aux consonnes de liaison.

1.2.2.1.4 Liaison et segments environnants

Nous examinerons à présent quelques cas particuliers concernant le segment phonique précédant immédiatement la consonne de liaison et celui la suivant immédiatement.

Intéressons-nous pour commencer au contexte phonique gauche immédiat, lorsque celui-ci est une voyelle. Son timbre peut être modifié dans certains cas de liaison. Évoquons tout d'abord le cas des liaisons en /n/, qui ne sont produites (hors pataquès) que pour des mots dont la finale phonique en isolation est une voyelle nasale. Ces liaisons,

si elles sont réalisées, peuvent avoir une influence sur cette voyelle les précédant, qui peut :

- rester nasalisée, comme dans le mot produit en isolation,
- se dénasaliser.

Selon les géolectes, cette voyelle précédant la liaison peut être ouverte, comme en français standard, ou fermée, comme chez les locuteurs du Fonds Jean Dumas (FJD) (cf *mon ami* [monami] dans l'exemple 107). Toutes les voyelles nasales peuvent-elles se dénasaliser dans ce contexte ? Selon Fouché (1956 : 435-436)[14] :

- /ɛ̃/ pourrait se dénasaliser en [ɛ] pour les trigrammes -‹ain›, -‹ein›, -‹ien› et -‹yen› (*certain auteur* [sɛʁtɛnotœʁ], *plein air* [plɛnɛʁ], *ancien état* [ɑ̃sjɛneta], *Moyen-Âge* [mwajɛnaʒ]), en [i] pour le digramme -‹in› (*divin Enfant* [divinɑ̃fɑ̃]) ou resterait nasalisé pour le digramme -‹un›[15] (*un ami* [ɛ̃nami]).
- /ɔ̃/ pourrait se dénasaliser en [ɔ] (*bon élève* [bɔnelɛv]).
- /ɑ̃/ resterait nasalisé (*en effet* [ɑ̃nefɛ]).

Les voyelles nasales précédant des liaisons en /n/ ne sont pas les seules voyelles pouvant être modifiées en contexte de liaison. Mentionnons à présent le cas de l'ouverture de la voyelle /e/ lors de liaisons en //ʁ// pour des adjectifs masculins singuliers en -‹er› ou -‹ier›. En isolation, la finale phonique de ces mots est la voyelle /e/. Si la liaison est réalisée, cette voyelle peut s'ouvrir et se réaliser [ɛ] (*premier homme* [pʁəmjɛʁɔm] ; Fouché, 1956 : 435).

Continuons avec un autre cas, celui des mots à consonne finale en isolation avec un schwa final potentiel. En contexte de liaison, un schwa peut apparaître lorsque le M1 est liaisonné afin d'éviter un groupe consonantique, ou si la consonne de liaison est de même nature que la consonne finale du M1, celle-ci peut être géminée (Bauche, 1920 : 57 ; Coquillon & Turcsan, 2012 : 121).

Enfin, penchons-nous sur un dernier cas de figure où le contexte phonique gauche, pour des lexèmes spécifiques, n'est pas réalisé en-dehors de la consonne de liaison. Viennent à l'esprit *nous*, *vous*, *ils* (exemple 19) pour lesquels seul un [z] peut apparaître devant voyelle, ou *est* (exemple 20), pour lequel seule la liaison [t] peut être produite dans le même contexte.

(19) *SGR : toute la nuit (il)s ont cheminé sans pouvoir trouver à loger.
%pho : tutə la nɥi zɔ̃ ʃəmine sɑ̃ puvwar truver a loʒe

[14] Tous les exemples proviennent de cette référence, la phonétique a été adaptée.

[15] Ce digramme était autrefois prononcé /œ̃/, mais l'opposition entre les phonèmes /œ̃/ et /ɛ̃/ a disparu en français standard contemporain au profit de /ɛ̃/.

(JD_a060_34, 43810, Virginie Granouillet)

(20) *SGR : la richesse n'est rien la beauté (es)t une rose.
%pho : la ʁiʃɛsə nɛ ʁjɛ̃ la bote tynə ʁozə
(Fonds Charles Joisten, M0996_SON76_19_A_01, Marcelle Sage)

Poursuivons en nous penchant sur le contexte phonique droit immédiat, en nous tournant vers le cas des liaisons devant les **‹h›-** improprement qualifiés d'***aspirés***. Il est plus adéquat d'utiliser l'appellation de ***‹h›- disjonctifs***. Il s'agit de certains ‹h› graphiques initiaux qui n'ont pas de correspondance phonique en synchronie et en français standard, mais qui, bien que l'initiale phonique du M2 soit vocalique, et dans des conditions par ailleurs tout à fait liaisonnantes, empêchent généralement que la liaison (et d'autre part l'élision) ait lieu. Comme nous le verrons en 1.2.3.1, ce sont des raisons diachroniques qui expliquent en règle générale que la liaison soit faite devant certains mots à ‹h› initiaux (on parle alors de ***‹h›- muets***) et pas devant d'autres[16]. Cependant, cela ne s'arrête pas à ce type de mots. Peuvent ne pas être liaisonnés (Fouché, 1956 : 439-440) :

- certains noms de lettres (*un a*, *un e*, *un f*, *un h*, *un i*, *un l*, *un m*, *un n*, *un o*, *un r*, *un s*, *un u*, *un x*, *un y*) et par conséquent, certains sigles (ex : *son* [on][17] *HLM* ; Tops 1, Jean Ferrat, *La Montagne*, 1965 [1964])
- certains noms de nombres cardinaux (*un 1*, *un 8*, *un 11*) et ordinaux (*un huitième*, *un onzième*)
- certains noms commençant par une glissante, notamment d'origine étrangère (ex : *les* [oz] *waters* (Tops 1, Chagrin d'Amour, *Chacun fait (c'qui lui plaît)*, 1982 [1981]) ; *mon* [on] *whisky* (Tops 1, Plastic Bertrand, *Ça plane pour moi*, 1978 [1977]) ; *des* [oz] *yéyés* (Tops 1, Les Vagabonds, *Le temps des yéyés*, 1990 [1989]))
- certaines entités nommées : anthroponymes (ex : *les* [oz] *O'Connelly* ; Tops 1, Michel Sardou, *Les lacs du Connemara*, 1981 [1981]) ou toponymes
- certaines interjections (ex : *des* [oz] ah, des [oz] oh ; Tops 1, Mylène Farmer, *Pourvu qu'elles soient douces*, 1988 [1988])

Ainsi, la réalisation de la consonne de liaison /n/ peut entraîner la dénasalisation de la voyelle la précédant (/ɛ̃/ en [ɛ] ou [i] et /ɔ̃/ en [ɔ]) ou non (/ɛ̃/ reste [ɛ̃], /ɔ̃/ reste [ɔ̃], /ɑ̃/ reste [ɑ̃]), quand la réalisation de la consonne de liaison //ʁ// peut entraîner l'ouverture de la voyelle /e/ la précédant en [ɛ]. La réalisation d'une liaison peut aussi être couplée d'une réalisation de schwa en amont ou une consonne finale

[16] Pour une liste détaillée de mots et d'entités nommées à ‹h›- disjonctifs, voir Fouché, 1956 : 251-265.
[17] Le système de codage des liaisons est présenté en 2.1.3.4.2.

peut être géminée lors de la réalisation d'une consonne de liaison de même nature. Par ailleurs, pour *nous*, *vous*, *ils* ou encore *est*, parfois seule la consonne de liaison indique la présence de ces lexèmes dans la chaîne parlée. Pour terminer, les liaisons ne se font ordinairement pas devant les mots à ‹h›- disjonctifs, les noms de lettres, les sigles, les noms de nombres cardinaux et ordinaux, certains noms commençant par une glissante (surtout d'origine étrangère), les entitées nommées ou encore les interjections.

Après nous être intéressés à la question du lien entre liaison et hiatus, avoir bien redistingué liaison et enchaînement, nous être penchés en profondeur sur les différentes consonnes de liaison, leur prégnance et leur influence sur des segments phoniques adjacents, il nous est toujours impossible de savoir quelle liaison peut être faite ou non. Comme le résume avantageusement Mallet (2008 : 55) : « [...] la condition phonologique est insuffisante pour rendre compte de la complexité du phénomène [de liaison]. [...] elle n'explique [...] pas ce qui la caractérise le plus : sa variation ». Dans la prochaine étape, nous verrons ce que le niveau morphologique a à nous apprendre.

1.2.2.2 Liaison et morphologie

1.2.2.2.1 Statut morphologique des consonnes de liaison

Si on adopte une approche morphologique, les consonnes de liaison peuvent être vues comme des morphèmes flexionnels, comme dans *petites* [z] *anecdotes* (PFC, Toulouse (France, Europe), 2006, 31agc1), ou comme appartenant au radical, comme dans *mauvais* [z] *état* (PFC, Bangui (République centrafricaine, Afrique subsaharienne), 2008, rcakn1) (Encrevé, 1983 : 40). Dans le premier cas, la liaison peut être vue comme un suffixe du mot liaisonnant ou comme un préfixe du mot liaisonné (Côté, 2011 : 8). Dans le second cas, il existe deux variantes supplétives, une en contexte de liaison (*mauvais* [z] *état*) et une hors contexte de liaison (*mauvais* [oz] *français* ; PFC, Saguenay (Québec, Canada), 2006, cqbma1) (Tranel, 1990 : 172), auquel cas il y a alternance entre radical long et radical court (Morin, 1986 : 170). Notons qu'en anglais, il n'existe qu'un terme, *stem*, quand en français on trouve deux termes, *radical* et *thème*. Les **thèmes** sont les « formes phonologiques associées à un même lexème » (Bonami & Boyé, 2003b : 104), apprises par l'auditeur. Roché (2010) propose de bien distinguer le thème du **radical**, ce dernier correspondant à « la chaîne segmentale à laquelle s'accroche l'affixe (flexionnel ou dérivationnel) » (Roché, 2010 : 99), dont se servent les locuteurs au cours d'opérations morphologiques. Les différents thèmes d'un même lexème sont réunis dans un **espace thématique**, qui « rassemble ses différents allomorphes et organise les rela-

tions qui les associent » (Roché, 2010 : 100). Les allomorphies conditionnées (Bonami & Boyé, 2003a : 151) sont enfin souvent mentionnées pour les adjectifs antéposés, avec d'un côté un allomorphe de liaison et de l'autre un allomorphe libre (Plénat, 2008).

Retenons donc que d'un point de vue morphologique, les liaisons sont des morphèmes flexionnels (suffixe du M1 ou préfixe du M2) ou appartiennent au radical (alternance radical long en liaison et radical court hors liaison). Nous porterons notre attention sur ces radicaux longs, entre autres, dans la sous-partie suivante consacrée à la notion de paradigme morphologique.

1.2.2.2.2 Liaison et paradigmes morphologiques

Les consonnes de liaison peuvent présenter des points communs avec d'autres consonnes utilisées dans le paradigme morphologique du mot liaisonnant, comme dans l'exemple suivant, souvent présenté : *petit* est prononcé en isolation [p(ə)ti], mais dans des cas de liaison comme *petit ami* [p(ə)titami], avec la forme féminine de l'adjectif *petite* [p(ə)tit] ou dans des mots dérivés tels que *petitesse* [p(ə)titɛs], un [t] est produit. Ici, le [t] est consonne de liaison, consonne flexionnelle et consonne dérivationnelle (Booij & De Jong, 1987 : 1005). Mais cela ne s'observe pas systématiquement : la consonne de liaison peut différer de la consonne flexionnelle ou dérivationnelle par le trait de voisement : [t] vs [d] (comme avec *grand* [gʁɑ̃] : *grand ami* [gʁɑ̃tami] vs *grande* [gʁɑ̃d] ou *grandeur* [gʁɑ̃dœʁ]) ou [z] vs [s] (comme avec *gros* [gʁo] : *gros arbre* [gʁozaʁbʁ] vs *grosse* [gʁos] ou *grosseur* [gʁosœʁ]) : une occlusive voisée marquant le féminin est non voisée en liaison et une fricative non voisée marquant le féminin est voisée en liaison. Notons que parfois l'alternance va plus loin qu'une présence vs absence de consonne finale entre deux allomorphes : nous avions vu en 1.2.2.1.4 que la réalisation d'une liaison pouvait influencer la prononciation de segments environnants tels que la voyelle précédente. Ces alternances seront présentées plus spécifiquement en 1.2.2.2.4.

Dans le paradigme morphologique d'un mot, la forme de liaison au masculin singulier peut encore être défective (Morin, 1992 : 245) : le mot n'est jamais utilisé en contexte de liaison (cf *brun*[18] dans le Tableau 1.1). Seules les données introspectives et les écrits oralisés pourront en présenter (Barreca, 2015). Ce cas de figure est extrêment fréquent car dans sa thèse de doctorat, Giulia Barreca ne trouve dans des corpus oraux (PFC et CFPP2000) que douze adjectifs masculins singuliers de liaison en position prénominale : *bon*, *petit*, *grand*, *saint*, *certain*, *mauvais*, *plein*, *ancien*, *bas*, *moyen*, *prochain* et

[18] En français standard contemporain, l'opposition /ɛ̃/~/œ̃/ tend à disparaître au profit d'un seul phonème, /ɛ̃/.

gros (Barreca, 2015 : 191). Selon Morin (1992 : 242) : « la forme liée du masculin doit être APPRISE indépendamment des autres pour pouvoir être utilisée normalement ». Étant donné que les auditeurs n'entendent pas ces formes, ils ne les produisent pas en tant que locuteurs et auront recours à des stratégies d'évitement : utiliser une autre construction syntaxique ou une forme de liaison au masculin singulier qu'ils maîtrisent (Morin, 1992 : 245). De ce fait, ces formes défectives perdurent en diachronie (Morin, 1992 : 245).

Faisons le bilan des six cas de figure qu'on rencontre par exemple pour les adjectifs prénominaux (exemples issus de Plénat, 2008) :

Finale en isolation	**Alternance**	**Nombre d'allomorphes**	**Exemple de lexème**	**Allomorphe libre masculin**	**Allomorphe prévocalique masculin**	**Allomorphe féminin**
vocalique	non	un	VRAI	vrai, vraie [vʁɛ] (*adjectif épicène à l'oral*)		
vocalique	oui	deux	PETIT	petit [pəti]	petit, petite [pətit]	
vocalique	oui	deux	BRUN	brun [bʁœ̃]	NA (*défectif*)	brune [bʁyn]
vocalique	oui	trois	GRAND	grand [gʁɑ̃]	grand [gʁɑ̃t]	grande [gʁɑ̃d]
consonantique	non	un	HONNÊTE	honnête [onɛt] (*adjectif épicène*)		
consonantique	oui	deux	COURT	court [kuʁ]		courte [kuʁt]

TABLEAU 1.1 : Paradigmes morphologiques des adjectifs prénominaux

Certaines consonnes de liaison se retrouvent de la sorte en flexion (marque du féminin) ou en dérivation, parfois avec une différence en terme de voisement. La plupart des adjectifs prénominaux masculins sont défectifs en contexte de liaison. Nous poursuivrons en nous interrogeant sur le degré de morphologisation des consonnes de liaison, qu'il s'agisse de morphémisation ou d'allomorphisation.

1.2.2.2.3 Morphologisation des consonnes de liaison

D'un point de vue fonctionnel, certaines consonnes de liaison ont pu se voir attribuer un rôle morphologique. Il s'agit des consonnes de liaison flexionnelles. Carton (1974 : 218) parle même de *morphème latent*, comme les consonnes de liaison sont parfois dites latentes. Le /z/ a pu être vu comme une marque de pluriel (Bauche, 1920 : 57 ; Frei, 1929 : 119 ; Delattre, 1947 : 149 ; Gadet, 1992 : 48 ; Bybee, 2005 : 28 ; Durand et al., 2011 : 115) ou comme une marque de première et deuxième personne (Morin & Kaye, 1982 : 325 ; Tranel, 1987 : 169 ; Desrochers, 1994 : 253), voire comme la marque des personnes d'allocution (cf 1.2.3.1.2.2) et le /t/ comme une marque de troisième personne (Morin & Kaye, 1982 : 325 ; Léon, 1984 : 5 ; Tranel, 1987 : 169). Sur cette lancée, j'avais pour ma part proposé de voir le /z/ comme une marque de participant au discours et le /t/ comme une marque de non-participant au discours (Coutanson,

2017 : 80). En ce qui concerne les marques de personne, elles ne s'apposent que sur des verbes conjugués[19], sans toutefois s'appliquer pour certains groupes verbaux, temps et modes ou encore personnes. Pour finir, si l'on s'intéresse à la liaison //ʁ//, celle-ci peut apparaître après les infinitifs en -‹er› et -‹ier›, mais pas après les participes passés de ces verbes, qui présentent cependant une forme en isolation similaire. C'est la raison pour laquelle les grammairiens du XVII^{ème} s. ont décidé de rétablir les -‹r› finaux des verbes en -‹ir› et -‹oir› (voir 1.2.3.1.1.3), « afin d'éviter les collisions homonymiques avec d'autres formes verbales, principalement avec les participes » (Englebert, 2009 : 130). Ceci pourrait mener à reconnaître //ʁ// comme marque d'infinitif (Coutanson, 2017 : 82). Notons toutefois que morphologiquement, //ʁ// marque aussi le futur ou le conditionnel en français contemporain (Marchello-Nizia et al., 2020 : 790). Dans ces derniers cas, //ʁ// suit la base mais précède les désinences de tiroir et de personne (Marchello-Nizia et al., 2020 : 790). Il existe un lien entre futur, conditionnel et infinitif puisque le //ʁ// marquant le futur et le conditionnel provient diachroniquement, pour la plupart des verbes, de l'infinitif (Marchello-Nizia et al., 2020 : 790, 798).

La question reste de savoir si ces consonnes de liaison, auxquelles on attribue un rôle morphologique, ont réellement été morphémisées par les locuteurs. Deux arguments peuvent être convoqués : l'existence de pataquès, qui indiqueraient une réanalyse morphologique des consonnes de liaison, matérialisée lorsque la liaison survient après des mots liaisonnants inattendus, ou une réalisation différentielle de liaisons selon des critères morphologiques. Nous aborderons le premier cas de figure en 1.3.5.2. Pour le second, il a pu être dit que le rôle fonctionnel d'une liaison peut favoriser sa réalisation. Par exemple pour le /z/ marque de pluriel, la liaison serait davantage réalisée après des noms pluriels alors qu'elle ne serait pas produite après des noms singuliers (Delattre, 1947 : 149). Ågren (1973 : 124) rappelle que « le fait même que la liaison du nom pluriel est facultative nous renseigne sur la valeur diminuée de cette fonction » et que « la notion de « pluriel » est la plupart du temps exprimée autrement et ailleurs dans le contexte », par exemple dans le déterminant du groupe nominal. À propos de ce dernier aspect, Léon (1992 : 243) parle de *marque morphologique redondante*. Laks (2005a : 104) émet également un doute concernant l'unicité du marquage morphologique de pluriel, qu'on rencontre à la fois dans des contextes invariables (*les amis*) et variables (*amis étrangers*). Que les liaisons soient vues comme une marque de pluriel, de personne (ou de (non-)participant au discours) ou d'infinitif, ces marques sont inlassablement redondantes. Enfin, leur présence uniquement en contexte de liaison (mais

[19]On parle parfois de ***verbe tensé***, ce qui exclut à la fois les infinitifs et les participes.

pas seulement pour //ʁ//) va à l'encontre d'une réelle morphologisation.

Dans deux cas précis, la littérature hésite à qualifier les consonnes de morphèmes et propose une allomorphisation des consonnes de liaison : il s'agit des occurrences qui se produisent devant les enclitiques (voir notamment Morin, 1986 : 190). Cela concerne en premier lieu les enclitiques *en* et *y* qui, postposés à un verbe à l'impératif, prennent les formes phoniques [zɑ̃] et [zi] (exemple 31). Cela concerne en second lieu les enclitiques *il*, *elle*, *on*, *ils* et *elles* qui, postposés à un verbe dans une construction interrogative, prennent les formes phoniques [ti(l)], [tɛl] et [tɔ̃] (exemple 32). *z-en* et *en*, *z-y* et *y*, *t-il* et *il*, *t-elle* et *elle*, *t-on* et *on*, *t-ils* et *ils* et pour finir *t-elles* et *elles* formeraient des couples d'allomorphes, chacun étant relié au morphème qui leur est propre, les variantes à initiale consonantique étant spécifiques aux formes enclitiques. Blanche-Benveniste (2010 : 62) fait le lien entre allomorphe long et accentuation : les enclitiques, postposés, sont accentués, contrairement à leurs équivalents antéposés, dépourvus d'accentuation.

En résumé, /z/ a pu être vu comme une marque de pluriel, de première ou de deuxième personne, /t/ comme une marque de troisième personne et //ʁ// pourrait être considéré comme une marque d'infinitif. Si ces consonnes ont réellement été morphémisées, on devrait trouver des pataquès ou une réalisation différentielle des liaisons selon des critères morphologiques. Par exemple, le /z/ marque de pluriel peut en effet favoriser la production de liaisons, sans toutefois la systématiser. De toute façon, le fait que seul le contexte de liaison présente ces marques n'illustre pas une morphémisation complète de ces consonnes, sans compter que leur valeur morphologique est redondante. Une analyse allomorphique des enclitiques à initiale vocalique est enfin parfois envisagée. Nous allons rester dans le domaine de la morphologie, en nous penchant sur le phénomène de neutralisation, dans un premier temps au niveau du genre du M1 et dans un second temps au niveau du nombre du M1.

1.2.2.2.4 Liaison et neutralisation morphologique

Commençons par préciser que certains linguistes préfèrent le terme de *syncrétisme*, et réservent le terme de *neutralisation* à la phonologie (voir Martinet, 1968).

Dans certains cas, le genre est neutralisé (cf Léon, 1992 : 243 et Blanche-Benveniste, 2010 : 62) au niveau du mot liaisonnant devant un nom liaisonné : il n'y a plus d'opposition masculin/féminin à l'oral pour le mot liaisonnant dès lors que le mot liaisonné est à initiale vocalique. Trois déterminants, les **déterminants unipossessifs**[20], pren-

[20] Il s'agit des « déterminants possessifs qui réfèrent aux trois personnes du singulier » (Chevalier &

dront une forme équivalente à celle du masculin devant des noms féminins : *mon*, *ton* et *son* + nom féminin à initiale vocalique (Bonami & Boyé, 2003a). En ancien français, c'était encore les formes féminines de ces déterminants qu'on trouvait devant initiale vocalique, élidées en *m'*, *t'*, *s'* (Chevalier & Fréchette, 2004 : 7 ; Marchello-Nizia et al., 2020 : 710). Par exemple, la forme *m'amie* de l'ancien français est devenue *mon amie* en français moderne (Chevalier & Fréchette, 2004 : 7). Les formes longues s'observent dès la fin du XII^{ème} s. en lorrain, sont courantes au XIV^{ème} s. en français mais ne se généralisent qu'au XV^{ème} s., voire plus tard selon les M2 en question (Marchello-Nizia et al., 2020 : 710). Au contraire, un déterminant (*cet*) et quelques adjectifs (*vieil*, *nouvel*, *bel*, *petit*...) prendront une forme équivalente à celle du féminin devant des noms masculins à initiale vocalique à l'oral (Bonami & Boyé, 2003a). Ainsi devant initiale vocalique, c'est la forme longue qui est mobilisée. Cela ne concerne que les variantes prénominales (ex : *un vieil/*vieux homme entra* vs *un homme *vieil/vieux entra* ; Bonami & Boyé, 2003a). De la sorte, Morin (1986 : 167) les considère comme des morphèmes liés, car leur forme phonique diffère de celle de leurs équivalents libres. Cela peut s'étendre à d'autres cas, étudiés par M. Plénat & C. Plénat dans un corpus du *Monde* (1991-2000) et sur internet (2011 : 304-305), telles les coordinations adjectivales en *et* antéposées à un nom à initiale vocalique dont le premier terme est BEAU, NOUVEAU, FOU, MOU ou VIEUX (ex : *son bel et puissant engin*), voire dans des coordinations adjectivales en *et* dont le premier terme correspond toujours aux mêmes lexèmes mais cette fois-ci antéposées à un nom à initiale consonantique (ex : *le bel et austère château de Charles Quint* ; Plénat & Plénat, 2011 : 211). Dans ce cas, ces auteurs notent que le second adjectif est souvent à initiale vocalique. Nous avons vu que les adjectifs prénominaux de liaison étaient rares dans les grands corpus oraux de français (1.2.2.2.2). Ces auteurs font donc le choix de travailler sur corpus écrit, mais doivent se limiter aux lexèmes susmentionnés car ce sont les seuls où la forme graphique standard diffère entre la forme liée et la forme libre, bien qu'ils mentionnent l'existence de graphies non standard utilisant la forme féminine en lieu et place de la forme de liaison au masculin singulier comme « *coucou titicia c mélissa ba ct pr te souhaiter un joyeuse et heureuse anniversaire sa fai koi davoir 18 ans allez bisou ma poule* » (Plénat & Plénat, 2011 : 306). Pour finir, n'oublions pas que parfois, aucune marque de genre ne subsiste dans l'énoncé à l'oral :

(21) Oui notre famille et et le et la famille de [mɔ̃nami] c'est très différente.
(PFC, Bamako (Mali, Afrique subsaharienne), 2006, maabh1).

Très brièvement, en ce qui concerne la neutralisation au niveau du nombre du M1,

Fréchette, 2004 : 7).

la réalisation d'une liaison après un verbe conjugué au présent de l'indicatif à la 3.SG peut parfois prêter à confusion, par rapport au même verbe à la 3.PL sans liaison (sinon il y aurait gémination de la consonne fixe finale). En effet, à ce temps et à ce mode en français, « les verbes forment des singuliers plus courts que les pluriels correspondants » (Blanche-Benveniste, 2010 : 50). La forme longue du radical en liaison peut coïncider avec la forme longue du radical à la 3.PL, comme dans l'exemple suivant proposé par M. Léon (1984 : 4) : *il bat un adversaire* vs *ils battent un adversaire*, prononcés tous deux [i(l)batɛ̃nadvɛʁsɛʁ].

Par conséquent, il peut y avoir neutralisation au niveau du genre pour quelques déterminants et quelques adjectifs en liaison, qui prendront la forme longue devant un nom à initiale vocalique. Cette neutralisation n'a pas toujours eu cours, comme nous avons pu le voir avec les déterminants unipossessifs féminins. Ces formes liées peuvent se retrouver dans les coordinations adjectivales en *et* dont le premier terme est *beau*, *nouveau*, *fou*, *mou* ou *vieux*, antéposées à un nom à initiale vocalique ou consonantique. La neutralisation du genre peut enfin être totale au sein du syntagme. Quant à la neutralisation au niveau du nombre, elle pourrait se produire dans de rares cas où la consonne de liaison de la 3.SG concorde avec la consonne fixe finale de la 3.PL.

Après nous être penchés sur les deux analyses morphologiques possibles des consonnes de liaisons, nous avons tenté d'appréhender notre phénomène d'étude en termes de paradigmes morphologiques pour les mots liaisonnants. Nous avons ensuite considéré une possible morphologisation des consonnes de liaison (morphémisation ou allomorphisation), mais nous avons vu que celle-ci était limitée, puis nous avons envisagé une possible neutralisation au niveau du genre et du nombre en liaison et à nouveau, celle-ci restait souvent limitée. Poursuivons à présent notre cheminement en regardant ce que le niveau syntaxique a à nous apprendre.

1.2.2.3 Liaison et syntaxe

1.2.2.3.1 Liaison et cohésion syntaxique

Chaque contexte de liaison est borné par un M1 et un M2. Ces mots peuvent être définis catégoriellement, c'est-à-dire selon leur catégorie morphosyntaxique. On peut alors regarder dans quelle mesure les combinaisons de ces catégories ou de ces mots sont liaisonnantes. Delattre (1947 : 148) indiquait déjà que « la liaison se fait dans la mesure où l'usage a consacré l'extrême étroitesse d'union de deux mots ou classes de mots ».

Lorsque le lien syntaxique entre M1 et M2 est fort, on parle parfois de *cohésion syntaxique* (Mallet, 2008 : 61) : une forte cohésion syntaxique entraînerait la réalisation d'un bon nombre de liaisons alors qu'une faible cohésion syntaxique ne favoriserait pas l'apparition de celles-ci. Les liaisons, ces anciennes consonnes finales (voir 1.2.3.1.1) auraient été préférablement conservées dans les contextes dits à forte cohésion syntaxique. Pour mesurer ce degré d'union, Delattre (1955 : 45) propose de tester la possibilité d'effectuer une pause à la frontière de mots étudiée, et fait donc ainsi appel à la prosodie. Ce n'est malheureusement pas suffisant, car Morin & Kaye (1982 : 301) soulignent le fait que des liaisons peuvent apparaître après des pauses. Qui plus est, dans l'exemple de Laks (2005b : 4) : *Des amis, j'en ai six à Lille* ([sisalil] et non [sizalil]), aucune pause n'est nécessaire et pourtant la liaison n'est pas réalisée, seulement un enchaînement. Bybee (2005 : 24) pour sa part propose de concevoir la cohésion syntaxique comme résultant de « la fréquence de co-occurrence » entre M1 et M2.

Mallet (2008 : 61) fait le pertinent constat suivant :

> La plupart [des] travaux fondent leur définition sur une analyse linéaire et catégorielle des éléments à lier. Et, dans le même temps, seuls deux éléments (M1 et M2) sont mentionnés : c'est dans la contiguïté que se joue le phénomène. Aussi, la combinaison syntagmatique complète de la phrase ou d'un groupe n'est-elle pas envisagée. La dimension syntaxique est, de ce fait, extrêmement réduite.

Pour aller dans ce sens, Laks (2005b : 7) met en avant les exemples suivants : *Il s'y est jeté corps͡et âme*, avec liaison ([kɔʁzeam]) vs **Il a sauvé son corps et son âme*, sans liaison ([kɔʁesɔ̃nam]). On est bien ici en présence des mêmes M1 et M2, *corps* et *et*, mais selon que le contexte de liaison apparaît dans la locution figée *corps et âme* ou en dehors de cette locution, la liaison sera ou non réalisée. Cela permet à Laks (2005b : 15-16) de dire que la liaison peut constituer un indice de figement pour les syntacticiens. Une analyse du phénomène de liaison dans le cadre des grammaires de dépendance, cadre théorique dans lequel pourrait s'ancrer le concept de « cohésion syntaxique », semble toujours aujourd'hui faire défaut.

Ainsi, l'importance du facteur syntaxique dans la réalisation d'une liaison pourrait sembler somme toute relative, puisque seules auraient une influence les catégories grammaticales des mots précédant et suivant le contexte de liaison. Pourtant, nous venons de voir que la prise en compte d'unités syntaxiques plus étendues pouvait être

pertinente pour expliquer la réalisation de certaines liaisons. La syntaxe reste néanmoins le facteur le plus déterminant pour évaluer le caractère liaisonnant du contexte en question, plus encore que la morphologie ou le lexique. Dans le point suivant, nous ferons un détour par une présentation des différents types de liaisons, étape théorique nécessaire avant de regarder quels contextes morphosyntaxiques sont liaisonnants.

1.2.2.3.2 Typologie des contextes de liaison : nommer et définir les types retenus

Jusqu'à aujourd'hui, Delattre (1947) reste une référence en ce qui concerne la classification des liaisons en trois grands types : *obligatoires*, *facultatives* et *interdites*. Selon lui, les **liaisons obligatoires** sont toujours faites, les **liaisons facultatives** sont « celles que le sujet parlant choisit de faire ou de ne pas faire selon le ton qu'il veut donner à son discours » et les **liaisons interdites** ne sont jamais faites (Delattre, 1955 : 48). Ces termes ont été repris dans de multiples travaux portant sur la liaison (comme Ågren, 1973 ; Ahmad, 1993 ; Chevrot et al., 2007 ; Wauquier, 2009 ; Adda-Decker et al., 2012), voire dans des ouvrages à destination d'apprenants du français.

Pourtant, depuis les années 1980, ces termes ont été critiqués, notamment par Encrevé (1983 : 47) :

> Délaissant le normatif pour le descriptif, il faudrait en toute rigueur rebaptiser selon le lexique de Labov les *liaisons obligatoires*, *facultatives* et *interdites* en *liaisons invariables* (ou catégoriques), *variables* et *erratiques*.

Pour le premier type, on rencontre aussi le terme de *liaisons invariantes*, pour le second type celui de *liaisons recommandées* et pour le troisième les termes de *liaisons inusitées* ou de *liaisons abusives* (Baude & Dugua, 2015 : 351). Pour ce dernier type, certains parlent encore de *liaisons non attestées*, de *liaisons aléatoires* ou de *liaisons impossibles* (Durand et al., 2011 : 105, 113).

Dans le cadre de cette thèse, j'utilise les termes de *liaisons invariables*, *liaisons variables* et *liaisons erratiques* d'une part et de *liaisons attestées* vs *non attestées* d'autre part. Je vais porter mon attention en premier lieu sur les liaisons invariables, les liaisons erratiques et les liaisons non attestées afin de borner le phénomène de liaison, avant de me tourner vers les liaisons variables.

Je considère les **liaisons invariables** comme des liaisons ayant une réalisation proche de 100%, plutôt que d'adopter des définitions telles que « [liaisons] qui se réalisent toujours quel que soit le contexte ou le locuteur » (Mallet, 2008 : 65). En effet, nous verrons en 1.2.3.3 qu'il existe des exceptions dans certaines variétés de français. Par ailleurs,

dans chaque corpus, on trouve des invariables non réalisées (voir les exemples 261 et 260), considérées comme erreurs de performance par certains (comme Morin & Kaye, 1982 : 295). Notons au passage que l'appellation *invariable* (ou *catégorique*) peut porter à confusion car celles-ci ne sont pas inéluctablement réalisées, mais je fais le choix de la conserver. Certains estiment que ces liaisons sont le socle commun de tous les francophones (Côté, 2017 : 15), de tous les registres ou encore de tous les sociolectes. Je préfère considérer des liaisons comme invariables pour un corpus donné. Par exemple, dans les corpus de professionnels de la parole publique, les contextes de liaison compris dans les invariables seront plus étendus. Il existe d'autre part une définition plus sociolinguistique que fréquentielle : « [l]es liaisons [invariables] sont celles dont l'absence provoquerait un jugement de non-grammaticalité » (Blanche-Benveniste, 2010 : 51). De la même façon, cette norme pourrait varier selon les locuteurs. Les contextes invariables sont des contextes à forte cohésion syntaxique. Laks (2005b : 164) indique que « [l]a liaison est catégorique lorsqu'elle se met en place depuis une catégorie morphosyntaxique seconde (auxiliaire, catégorie régie etc.) vers une catégorie principale (tête, majeure, etc.) ». Le contre-exemple des enclitiques postposés aux verbes, où la liaison, bien que généralement absente dans les corpus oraux contemporains, est invariablement réalisée, n'est pas évoqué.

La dernière catégorie, les « interdites » de Delattre, manque de clarté : Delattre lui-même (1956 : 52) semble considérer ces liaisons comme plus ou moins interdites, puisqu'il mentionne des « liaisons complètement interdites » mais indique pour d'autres qu'« il faut admettre que l'interdiction n'est pas aussi catégorique »[21]. Les secondes pourraient de la sorte être produites alors que les premières ne seraient jamais observées. De ce fait, Delattre nous montre que cette dernière catégorie regroupe deux types de liaisons : des liaisons qui, de façon prescriptive, ne devraient pas être faites et ne sont pas produites par les locuteurs, et des liaisons que la prescription rejette mais qui pourraient être observées, en faible nombre. Comme nous l'avons vu plus haut, Encrevé (1983 : 50) propose de remplacer le terme de *liaisons interdites* par *liaisons erratiques*. De la même façon, Encrevé ne prend en compte qu'une des facettes, cette fois-ci la seconde. Laks saisit bien la différence en mélangeant les typologies de Delattre et d'Encrevé et en distinguant les obligatoires, les facultatives, les erratiques et les interdites (Laks, 2008 : 241). De mon côté, pour ne pas qu'il y ait de lacune, je considère ici comme **liaisons erratiques** les liaisons très rares (des liaisons ayant une réalisation proche de 0%, toujours pour un corpus donné) et comme **liaisons non attestées** des liaisons qui ne se

[21] Cela a aussi été noté par Pustka (2017 : 200), qui y voit une simplification didactique.

rencontrent jamais en corpus (à nouveau, pour un corpus donné) : les contextes sont connus pour ne pas être productifs dans le corpus étudié.

Sur cette lancée, les **liaisons attestées**, au contraire, sont les liaisons qui s'observent en corpus (qu'elles soient invariables, variables ou erratiques) : les contextes sont connus pour être productifs dans le corpus étudié.

Par défaut, les **liaisons variables** correspondent enfin aux liaisons qui ne rentrent ni dans la catégorie des liaisons invariables, ni dans la catégorie des liaisons erratiques, avec un taux de réalisation très fluctuant selon les contextes grammaticaux et lexicaux concernés et selon les corpus d'étude. Delattre (1955 : 43) note bien que les liaisons facultatives (donc variables) confrontent le locuteur à un choix : faire ou ne pas faire la liaison. C'est pour cela que les linguistes s'y sont tout particulièrement intéressés. Contrairement aux liaisons invariables, les contextes de liaisons variables seraient des contextes à moindre cohésion syntaxique. Laks (2005b : 164) indique cette fois-ci que ces contextes concerneraient des catégories morphosyntaxiques secondes postposées à des catégories principales.

Nous avons vu qu'il était préférable de remplacer la typologie des contextes de liaison en obligatoires, facultatives et interdites, prescriptive, par une typologie des contextes de liaison descriptive en liaisons attestées d'une part (invariables, variables et erratiques) et non attestées d'autre part. Une typologie par corpus me semble préférable. Nous en venons à la question suivante : a-t-on réellement besoin d'une subdivision en types de liaisons ? N'est-on pas plutôt en présence d'un continuum ? Celui-ci irait des liaisons invariables, globalement toutes réalisées hormis quelques exceptions, aux liaisons variables, par essence ayant des taux de réalisation très fluctuants, puis aux liaisons erratiques, très faiblement réalisées, avant d'atteindre les liaisons non attestées, absentes en corpus (Laks, 2008 : 30 envisage les choses de la même façon). Il est temps de se pencher sur les contextes compris dans chaque catégorie, ainsi que sur les révisions qui ont été faites de la typologie au fil des études.

1.2.2.3.3 Typologie des contextes de liaison : évolution des contextes retenus pour chaque type

Commençons par redonner le tableau récapitulatif de Delattre (figure 1.7 ; 1947 : 152), qui reste le point de départ de bon nombre des propositions de classification qui suivront. Dans le même article, le linguiste présente un tableau détaillé de cinq pages qui se veut le plus explicite possible (1947 : 153-157). Depuis lors, les contextes sont souvent regroupés sous forme de listes, comme le rapporte Mallet (2008 : 64). Dans

ses deux articles suivants, Delattre s'intéresse plus particulièrement aux liaisons facultatives, notamment aux facteurs influençant leur réalisation (Delattre, 1955), puis se concentre sur quelques contextes morphosyntaxiques de liaison spécifiques (Delattre, 1956).

TABLEAU SIMPLIFIÉ

	OBLIGATOIRES	FACULTATIVES	INTERDITES
NOM	*déterminatif* + {*nom* / *pronom* / *adjectif*} vos enfants deux autres un ancien ami	*nom pluriel* + des soldats anglais ses plans ont réussi	*nom singulier* + un soldat anglais son plan a réussi
VERBE	*pronom personnel* + *verbe* ils ont compris nous en avons *verbe* + *pronom personnel* ont-ils compris allons-y	*verbe* + je vais essayer j'avais entendu dire vous êtes invité il commençait à lire	
INVARIABLES	*invariables monosyllabiques* + en une journée très intéressant	*invariables polysyllabiques* + pendant un jour toujours utile	*et* + et on l'a fait
SPÉCIALES	*formes figées* comment allez-vous les Etats-Unis accent aigu tout à coup de temps en temps		*h aspiré* des héros en haut + *un, huit, onze et dérivés* la cent huitième en onze jours

FIGURE 1.7 : Tableau simplifié de la classification de Delattre (1947 : 152)

Le point faible de cette proposition est qu'elle est issue d'un travail par introspection. Rappelons qu'à l'époque de cette publication, il n'existait pas encore de corpus oral pour l'étude de la liaison. Ces résultats introspectifs peuvent aujourd'hui être plus ou moins corroborés par des observations sur corpus. Mallet (2008 : 67-80) propose un compte-rendu des diverses révisions de la proposition initiale de Delattre (1947) :

- des contextes indiqués comme obligatoires par Delattre seraient à considérer plutôt comme variables :
  - verbe impersonnel + M2[22] (cf Ahmad, 1993 : 108-109 sur corpus[23], Durand

[22]Pustka (2017 : 200) y voit une simplification didactique, puisque Delattre nuance ses propos dans un article ultérieur (Delattre, 1956 : 50).

[23]Le corpus d'Ahmad (1993 : 85-87) comprend trois sous-corpus : un corpus radiophonique des années 1970 dans un registre soutenu, principalement monologal; un corpus de conversations spontanées des hautes classes socio-économiques parisiennes des années 1970 (volume le plus important) et un corpus de conversations spontanées entre pairs datant de 1985.

& Lyche, 2008 : 19 dans le corpus PFC et Pustka, 2017 : 201 dans un corpus de livres audio pour enfants)

- adjectif + nom (cf Morin & Kaye, 1982 : 295 sur collection d'exemples et/ou par introspection ; Durand & Lyche, 2008 : 19 et Barreca, 2015 dans le corpus PFC)
- préposition monosyllabique + M2 (cf Malécot, 1975 : 164 sur corpus[24] ; Morin & Kaye, 1982 : 295 sur collection d'exemples et/ou par introspection ; De Jong, 1994 dans le corpus des ESLO (ESLO1) ; Durand & Lyche, 2008 : 17 et Mallet, 2008 : 277 dans le corpus PFC ; Laks, 2008 : 261 dans le corpus Hommes politiques (HPOL) ainsi qu'Adda-Decker et al., 2012 : 123 dans le corpus NCCFr)
- adverbe monosyllabique + M2 (cf Ågren, 1973 : 135 dans un corpus radiophonique enregistré dans les années 1960 ; Malécot, 1975 : 164 sur corpus ; Encrevé, 1983 : 52 dans un corpus de personnalités politiques ; De Jong, 1994 dans le corpus des ESLO (ESLO1) ; Ahmad, 1993 : 110, 122-123 sur corpus ; Mallet, 2008 : 281 dans le corpus PFC , Laks, 2008 : 261 dans le corpus HPOL et Pustka, 2017 : 204 dans un corpus de livres audio pour enfants)

Pourtant, comme le relève Pustka (2017 : 200, 204), Delattre est moins catégorique dans son article de 1956 : cette fois-ci, il présente l'exemple *c'est impossible* comme liant « presque toujours » et indique qu'après les prépositions et adverbes monosyllabiques : « [e]n gros, on peut considérer cette liaison comme obligatoire, et c'est ainsi qu'il faut la donner dans l'enseignement pratique. Mais si l'on regarde de plus près, on y découvre des tendances facultatives », avant de mentionner les cas de préposition monosyllabique + déterminant (*dans un an*), celui d'adverbe monosyllabique + participe passé (*trop amusé*), celui de *pas/plus* + M2 (*pas important, plus un sou*) et celui de *vers* + M2 (*vers elle* ; Delattre, 1956 : 50-51).

- des contextes indiqués comme facultatifs par Delattre sembleraient ne pas être attestés en corpus :
  - nom pluriel + verbe (cf Durand & Lyche, 2008 : 21 dans la tâche de lecture du corpus PFC)
  - préposition polysyllabique + M2 dans la parole ordinaire (notamment dans PFC ; Durand et al., 2011 : 117-118 mais aussi dans le corpus de Malécot,

[24] Le corpus de Malécot (1975) est le sous-corpus repris par Ahmad (1993), décrit dans la note de bas de p.23 : « corpus de conversations spontanées des hautes classes socio-économiques parisiennes des années 1970 ».

1975 : 104).

- aucun contexte indiqué comme interdit par Delattre n'a été contesté.

La question est aussi de savoir si le changement de catégorie doit affecter tous les items lexicaux. Par exemple, malgré le reclassement du contexte adjectif + nom en contexte variable, il est parfois proposé de conserver *petit* + nom parmi les contextes de liaisons invariables (Mallet, 2008 : 73), bien que Mallet elle-même (2008 : 235) indique que les liaisons n'y sont réalisées qu'à 75% (36/48). C'est également le cas, pour le contexte préposition monosyllabique + M2 recatégorisé comme un contexte variable, de la préposition *en*, qui ferait exception et resterait invariable (Mallet, 2008 : 80 ; Côté, 2017 : 15), quoique huit occurrences puissent être observées dans PFC (Côté, 2017 : 16), cette liaison n'étant ainsi pas systématiquement faite.

Concernant les modifications de la typologie initiale, comme le résume Encrevé (1983 : 48), « [s]oit [...] la position de Delattre [a] été descriptivement inexacte, soit [...] la situation [a] évolué ». Rappelons que ce linguiste ne travaillait pas à partir de corpus. Ce manque de données empiriques a pu l'amener à des inexactitudes lors de l'établissement de sa typologie. D'autre part, Delattre (1956 : 49) prenait comme référence « la conversation naturelle *de la classe cultivée* » du début du XX^{ème} siècle. La typologie proposée pourrait être caractéristique de ce sociolecte, tout comme la langue a pu évoluer en diachronie. La lacune principale de cette proposition est que cet auteur manquait cruellement des abondantes données empiriques sur la liaison actuellement à notre disposition, que ce soit en diastratie, en diaphasie ou encore en diatopie, nécessaires pour généraliser ses observations par introspection. En catégorisant des contextes comme invariables, variables, erratiques et non attestés propres à un unique corpus d'étude, comme proposé en 1.2.2.3.2, nous évitons une évolution du contenu des catégories au fil des nouvelles observations sur corpus concernant la liaison. Notamment, certains corpus relevant de la distance (Koch & Oesterreicher, 2001) pourraient admettre davantage de contextes de liaisons invariables.

La proposition de classification des contextes de liaison de Delattre (1947), introspective, a été révisée grâce à d'autres données introspectives, à des collections d'exemples, à des tâches de lecture et surtout grâce à l'utilisation de corpus oraux pour l'étude de la liaison. Reste à savoir si un contexte lexical spécifique peut être catégorisé séparément des autres membres du même contexte grammatical. Enfin, il a pu se produire une évolution diachronique en français demandant une recatégorisation de certains contextes ou la méthodologie d'analyse et les caractéristiques sociolinguistiques

du français de référence de Delattre ont pu l'empêcher de présenter une classification générale.

Dans cette partie, nous nous sommes intéressés au concept de cohésion syntaxique, en constatant que bien souvent seules les catégories morphosyntaxiques des mots précédent et suivant sont prises en compte et non des unités syntaxiques élargies, plus pertinentes dans certains cas. Je me suis ensuite rangée du côté d'une terminologie typologique descriptive des contextes de liaison où j'ai circonscrit et défini quatre catégories. Finalement, nous avons revu l'inventaire des contextes de liaison compris dans chaque type et nous nous sommes remémoré les révisions qui avaient déjà été proposées. Nous allons désormais interroger le lien qui existe entre liaison et lexique.

1.2.2.4 Liaison et lexique

Le lien entre liaison et lexique est évident, puisque nous savons dorénavant que la liaison apparaît entre deux mots, un M1 et un M2. Pour rappel, dans de très rares cas, le M1 n'est pas produit et seule la consonne de liaison indique sa présence (voir 1.2.2.1.4).

Comme nous l'avions vu en 1.2.2.2.1, les consonnes de liaisons peuvent être morphologiques ou appartiennent au radical (Encrevé, 1983 : 40), ces dernières étant ainsi des consonnes lexicales. De la sorte, de nombreuses consonnes de liaison sont liées à un terme lexical particulier.
Par exemple, en ce qui concerne la nature des consonnes de liaison, les moins fréquentes sont celles qui sont les plus liées à des lexèmes particuliers :

- /p/ s'observe uniquement après les adverbes *trop* et *beaucoup*, ce que confirment les données de PFC[25].
- //ʁ// n'apparaît qu'après des infinitifs et des adjectifs en -‹er›. Dans PFC seuls *se débrouiller*, *premier* (principalement) et *dernier* ont été notés dans les conversations[25].
- /n/ n'est produit qu'après *ancien*, *aucun*, *b(i)en*, *bon*, *certain*, *en*, *grand*[26], *mon*, *non*, *on*, *plein*, *rien*, *son*, *ton*, *un* (et nombres finissant en *un*) dans les conversations de PFC[25].

Pour ce qui est des liaisons non enchaînées, Laks (2008 : 248) a remarqué que pour

[25]Ces données ont été obtenues depuis le site https ://research.projet-pfc.net/ (extraction le 13 janvier 2023).

[26]Avec pataquès en [*n].

HPOL1 (le premier volet du corpus HPOL[27]), ce type de liaisons, bien qu'en faible nombre, ne touche qu'un petit nombre d'éléments lexicaux seulement : 17 pour le /z/ et 13 pour le /t/. Toutefois, comme l'explique Encrevé (1983 : 51), pour un même lexème, la liaison peut être réalisée de manière enchaînée comme non enchaînée au sein du même discours d'un locuteur donné.

Par ailleurs, le phénomène de liaison relevant d'un sandhi externe, il contribue à effacer les frontières de mots (Blanche-Benveniste, 2010 : 53), ce qui peut conduire à des lexicalisations ou relexicalisations. Laks (2014) présente les liaisons variables comme des liaisons qui ne seraient pas lexicalement stabilisées. Sur la même lancée, on pourrait envisager les liaisons invariables comme des liaisons davantage stabilisées, même s'il est essentiel de rappeler que toutes n'apparaissent qu'en contexte de liaison, à la différence de phénomènes diachroniques de métanalyses où certaines consonnes de liaisons ont été agglutinées au mot suivant (notamment pour *nombril*, exemple présenté en 1.2.3.1.1.3 ou encore pour *zieuter/zyeuter* 'regarder' avec agglutination du [z] de liaison sur *yeux* couplée d'une dérivation nom→verbe ; CNRTL, 2012n) et donc relexicalisées. On parle ici de lexicalisation car la consonne passe du statut de consonne de liaison à celui de consonne fixe, et que le mot n'est plus obligé de suivre le ou les M1 qui déclenchaient cette liaison. En 1.3.6.3, nous verrons qu'en français louisianais le phénomène d'agglutination a été encore plus massif, mais que la lexicalisation des consonnes de liaison varie selon les locuteurs. Les créoles à base française présentent eux aussi de nombreuses agglutinations lexicalisées, de façon beaucoup plus systématique.

Enfin, la liaison apparaît parfois dans des locutions figées (ex : *de temps en temps*) ou collocations quasi-figées (ex : *accent aigu*) qui relèvent de la phraséologie, et que nous avions déjà abordées dans 1.2.2.3.1. Ici, c'est un fort lien syntaxique, un tout sémantique et lexical qui explique la réalisation des liaisons. Laks (2005a : 115-116) rappelle cependant que cette dernière n'est pas totalement prévisible et ne présente pas de stabilité diachronique.

Nous avons donc vu que bon nombre de consonnes de liaison sont lexicales. J'ai aussi mentionné le fait que les consonnes de liaison /p/, //ʁ// et /n/ étaient les plus liées à quelques items lexicaux spécifiques. C'est également le cas des liaisons non enchaînées. Nous avons aussi abordé le cas de liaisons ayant été lexicalisées en diachronie et pour finir, celui de liaisons apparaissant dans une unité phraséologique. Il ne nous reste plus

[27] Le corpus est brièvement présenté en 1.2.3.1.1.4.

qu'à aborder le phénomène de liaison sous un angle sémantique.

1.2.2.5 Liaison et sémantique

À l'oral, la réalisation ou non-réalisation d'une liaison peut permettre des oppositions de sens, comme dans les exemples suivants proposés par Delattre (1947 : 151) :

(22) *Coupez-vous-en dix !* [kupevuzãdis] vs *Coupez-vous en dix !* [kupevuãdis]

(23) *Les auteurs* [lezotœʁ] vs *Les hauteurs* [leotœʁ].

Léon (1992 : 243) parle du *rôle distinctif* de la liaison dans ces paires minimales.

Au contraire, parfois certains énoncés peuvent se révéler équivoques, comme dans l'exemple suivant relevé à la radio par Blanche-Benveniste (2010 : 53), dans lequel le locuteur précise sa pensée afin que les auditeurs ne méprennent pas *qu'on naît tous* pour *qu'on est tous* :

(24) *Je pense qu'on naît tous - verbe naître - avec cette passion.*

Ici, un double sens peut être détecté entre un énoncé sans et un énoncé avec liaison : il y a homophonie et homonymie. Ces cas semblent pourtant très rares à l'oral, où le contexte permet souvent de désambiguïser l'énoncé en question. Ceci peut mener au calembour : Encrevé (1983 : 64) relève même chez un dirigeant de parti, lors de la campagne pour les élections législatives de 1978, la réparation *Quand Monsieur Mitterrand était ministre, et Dieu sait qu'il l'a beaucoup͡ été, euh beaucoup͡ été…*, où à une liaison réalisée avec enchaînement le locuteur préfère une liaison sans enchaînement pour éviter un double sens humoristique (*beaucoup été* vs *beaucoup pété*).

Enfin, nous avions vu en 1.2.2.4 que certaines liaisons pouvaient être relexicalisées avec le M2. J'avais cité l'exemple de *nombril*, pour lequel l'ancien terme n'a pas subsisté en français standard. Mais il existe aussi des cas de relexicalisation où les deux termes coexistent, celui avec une ex-consonne de liaison devenue une consonne fixe initiale prenant une nouvelle signification. C'est par exemple le cas de *zoreille*[28], provenant diachroniquement d'une agglutination d'un [z] de liaison sur le mot *oreilles*, et qui correspond en français calédonien[29] à l'ethnonyme « Métropolitain : originaire de la Métropole (France métropolitaine) » (ex : *Guillaume Peyre est un zoreille […]*) ou à un adjectif qui signifie : « [r]elatif à la Métropole » (ex : *le nouveau lotissement rose-orange, couleur zoreille* ; cf Pauleau, 2022).

[28] Ce mot présente plusieurs graphies.

[29] On le retrouve aussi à la Réunion, à l'Île Maurice, aux Antilles et en Polynésie (Pauleau, 2022).

De la sorte, la liaison pourrait jouer un rôle sémantique distinctif, sa réalisation ou non-réalisation renvoyant à deux énoncés de sens bien différents ou réciproquement, elle pourrait jouer un rôle neutralisateur, créant une homophonie entre un énoncé avec et un énoncé sans liaison, homonymes, donnant parfois naissance à un calembour. Ces deux phénomènes restent cependant épisodiques à l'oral tout comme la reformulation dans le second cas. Pour terminer, la relexicalisation d'une consonne de liaison peut entraîner un changement sémantique.

Nous avons abordé la liaison sous l'angle des différents niveaux linguistiques : phonologie (et phonétique), morphologie, syntaxe, lexique et finalement sémantique. À présent, la sociolinguistique a encore beaucoup à nous apprendre sur ce phénomène, et nous commencerons avec la dimension diachronique, avec une orientation davantage en linguistique diachronique.

1.2.3 La liaison : facteurs sociolinguistiques

1.2.3.1 Liaison et variation diachronique

1.2.3.1.1 La liaison en phonétique diachronique

1.2.3.1.1.1 Données en phonétique diachronique

Lorsqu'on travaille sur un phénomène oral comme la liaison, le mieux est d'avoir accès à des documents oraux. Cependant, l'enregistrement sonore remonte seulement à un peu plus d'un siècle, ce qui n'est pas suffisant pour appréhender l'évolution diachronique de notre phénomène. Pour connaître les états de langue plus anciens, les diachroniciens qui s'intéressent à la phonétique compilent des observations issues de différentes sources :

- de grammaires du français. On trouve des grammaires descriptives conçues par des Anglais dès la fin du XIII^{ème} s., puis des grammaires normatives à partir du XVII^{ème} s. (Englebert, 2009).
- de textes poétiques, en particulier en examinant les assonances et les rimes qu'on y trouve (afin de voir dans quelle mesure de mêmes graphies sont homophones) ou la métrique de ces textes (afin de se prononcer en faveur de la présence d'une diphtongue ou au contraire d'une monophtongue pour un digramme vocalique, ou encore en faveur d'une diérèse ou d'une synérèse à nouveau pour un digramme vocalique ; Englebert, 2009).

- plus généralement de tout texte, en s'attardant sur les hésitations graphiques. En effet, la correspondance graphie/phonie qu'on observe à l'émergence de l'ancien français ne fonctionne déjà plus dès la fin de ce chronolecte (Englebert, 2009). L'apparition de nouveaux phonèmes en français amène le système graphique de la langue à s'adapter à un système déjà existant, le système graphique latin (Englebert, 2009). Des graphèmes qui deviennent interchangeables peuvent rendre compte qu'ils représentent désormais un même son (Englebert, 2009).
- du champ de la grammaire comparée (par exemple, la prononciation des mots anglais d'origine normande donne des renseignements sur la prononciation de ce dialecte d'oïl à l'époque où ces mots sont arrivés dans la langue anglaise ; Englebert, 2009).

Ainsi, les enregistrements sonores remontant au maximum à un peu plus d'un siècle, un diachronicien s'intéressant à la phonétique recourra à des grammaires (descriptives ou normatives), des textes poétiques (étude des assonances, des rimes ou de la métrique), examinera les hésitations graphiques des textes et fera appel à la grammaire comparée. Dans le point suivant, je présenterai les dynamiques principales à l'œuvre dans l'évolution diachronique de la liaison.

1.2.3.1.1.2 Évolution diachronique de la liaison : dynamiques contradictoires

À l'origine, les liaisons étaient des consonnes finales, c'est-à-dire prononcées quel que soit le contexte : que le mot soit en isolation, qu'il se trouve devant initiale consonantique ou vocalique. Elles ont progressivement cessées d'être prononcées, tout d'abord devant initiale consonantique, puis à la pause à partir de la fin du XV^{ème} s. (Marchello-Nizia et al., 2020 : 399). Enfin, c'est dès le XVI^{ème} s. qu'on commence à remarquer des liaisons non réalisées, dans des contextes aujourd'hui considérés comme variables ou non attestés (Marchello-Nizia et al., 2020 : 475). On pourrait ainsi voir le phénomène de liaison comme un reliquat diachronique visible en synchronie. Cette conclusion est cependant trop hâtive, l'évolution diachronique de la liaison étant plus complexe. Marchello-Nizia et al. (2020 : 483-485) avancent quatre dynamiques contradictoires : une dynamique de syllabation ouverte, une dynamique de marquage morphologique, une dynamique d'enchaînement généralisé et une dynamique graphique. Arrêtons-nous sur chacune de ces dynamiques.

Commençons par la dynamique de syllabation ouverte. En latin, les monosyllabes sont des **oxytons** (mots accentués sur la dernière syllabe), les bisyllabes des **paroxy-**

tons (mots accentués sur la **pénultième**, c'est-à-dire l'avant-dernière syllabe) et les mots de plus de deux syllabes sont soit des paroxytons, soit des **proparoxytons** (mots accentués sur l'**antépénultième**, c'est-à-dire l'avant-avant-dernière syllabe ; Laborderie, 1994). Une forte accentuation de la **tonique** latine, c'est-à-dire de la syllabe accentuée, a affaibli les fins de mots, les faisant tout d'abord devenir paroxytoniques ou oxytoniques pour arriver exclusivement à des oxytons (Thibault, Année universitaire 2021-2022). Le français « est marqué par une dynamique syllabique qui privilégie une syllabation ouverte, spécialement pour la finale, quelle que soit l'unité considérée (mot, morphème, groupe intonatif etc.) », une tendance qui se profilait déjà dans les langues indo-européennes archaïques et en latin (Marchello-Nizia et al., 2020 : 483). Les consonnes finales, à la fois en finale de mot et de syllabe, se trouvaient dans une position faible qui a favorisé leur chute (Laborderie, 1994). Les consonnes s'amuïssent dans un premier temps : par exemple, dès le latin mérovingien, les -‹t› lexicaux latins finaux sont parfois notés -‹d›, ce qui laisse penser qu'un voisement s'est produit, quand certains ne sont déjà plus notés (Russo, 2014 ; voir également le sort de la consonne secondaire [t] des participes passés en 1.2.3.1.1.3, qui se voise à l'intervocalique, se spirantise, devient finale avant de se dévoiser et de s'amuïr).

En ce qui concerne la dynamique de marquage morphologique, lorsqu'il est réalisé sous forme de consonnes finales, ces dernières auront tendance à être préservées dans un premier temps, même si elles ne subsistent plus qu'en liaison aujourd'hui (Marchello-Nizia et al., 2020 : 484). C'est ce que nous verrons pour les marques de nombre -[z] et de personne -[z] et -[t] dans le point suivant (1.2.3.1.1.3). La réalisation de consonnes flexionnelles de liaison peut être conscientisée car il s'agit d'un « savoir savant sur la langue » (Marchello-Nizia et al., 2020 : 487).

Venons-en à la dynamique d'enchaînement généralisé. Les consonnes finales se sont uniquement maintenues entre des mots appartenant à la même unité prosodique (Gess et al., 2012 : 7). En effet, l'accentuation de mot du latin a été abandonnée au profit d'une accentuation de syntagme en français entre le XII^{ème} et le XIII^{ème} s., ce qui a favorisé des comportements similaires entre consonnes finales à l'intérieur du syntagme et consonnes internes (effacement devant consonne, maintien et enchaînement à l'intervocalique ; Marchello-Nizia et al., 2020 : 470, 484, 2002).

Il ne nous reste plus qu'à aborder la dynamique graphique. Ce qui a été considéré comme contexte de liaison selon les époques est lié à l'évolution de l'orthographe française. Il faut savoir qu'en ancien français, l'orthographe était beaucoup plus phonétique qu'en français contemporain (Perret, 1998). Dès le moyen français, l'orthographe se complexifie, facilitant parfois la lecture avec l'utilisation en particulier de graphies

étymologisantes ayant une fonction discriminante (par exemple, l'ajout d'un ‹d› à *pied* (latin (lat.) PEDEM) permet de le distinguer de *pie* (l'oiseau)), facilitant une identification du mot par appréhension globale (Perret, 1998). Le ‹d› de *pied* a été rétabli à l'écrit, même s'il n'avait pas été conservé à l'oral, ce qui n'empêche pourtant pas de dire aujourd'hui un *pied-à-terre* [pjetatɛʁ]. La liaison de ce nom composé n'est attestée que depuis le XVIII^{ème} s. (CNRTL, 2012k), par conséquent la réalisation de certaines liaisons est liée à l'évolution de l'orthographe française. Les graphies étymologisantes ou pseudo-étymologiques foisonnent avec le développement de l'imprimerie en Europe (Marchello-Nizia et al., 2020 : 485). Dès le XVII^{ème} s., dans le cadre d'une tentative d'établissement d'une norme graphique pour le français, certaines de ces graphies sont supprimées (*nud* > *nu*, *bled* > *blé*… ; Perret, 1998 ; Marchello-Nizia et al., 2020 : 485). Il ne vient pas à l'idée que ces derniers mots puissent liaisonner, bien qu'à une époque donnée, ils aient porté une consonne finale. C'est tout le problème de se fonder principalement sur la consonne graphique finale des mots en français contemporain afin de déterminer s'ils peuvent liaisonner. Les consonnes finales ne sont pas rétablies seulement à l'écrit mais également à l'oral, sous l'influence de la lecture à voix haute, des récitations et des dictées orales qui se développent dans l'enseignement scolaire, par « effet Buben » (Marchello-Nizia et al., 2020 : 485-486, 488 ; Buben, 1935). Marchello-Nizia et al. (2020 : 488) indiquent que « [j]usqu'au milieu du 19e siècle, l'usage ordinaire, mais aussi l'usage soutenu, sont […] relativement pauvres en liaisons variables ». À partir de cette époque, la réalisation de nombreuses liaisons variables devient un marqueur d'éducation, de culture et de position sociale (phénomène renforcé à la fin du XX^{ème} s. avec l'apparition des liaisons sans enchaînement), quand d'un point de vue prescriptif, les pataquès viennent à être considérés comme fautifs (Marchello-Nizia et al., 2020 : 486, 489).

On pourrait ajouter à cela deux dernières dynamiques qui interagissent de diverses façons avec les autres, la première étant une dynamique fréquentielle. Yves-Charles Morin (2005 : 18-19) rejoint Mildred Katherine Pope (1952) selon laquelle l'émergence de la liaison serait due à un processus analogique, soit la généralisation progressive de la variante préconsonantique au contexte prépausal puis prévocalique. Selon Bybee (2002 : 218-219), c'est la fréquence plus élevée des contextes préconsonantaux dans la langue qui aurait initié cette évolution diachronique. Toujours d'après Bybee (2005 : 30), les liaisons apparaissant dans une construction peu fréquente seraient les premières à disparaître, alors que celles qui seraient réalisées au sein de constructions fréquentes seraient productives et résisteraient plus longtemps.

Une ultime dynamique à l'œuvre en français à l'origine du phénomène de liaison

est la dynamique analogique. Celle-ci interagit notamment avec la dynamique morphologique (nous verrons par exemple en 1.2.3.1.2.2 que la marque de personne [z] s'est étendue à la 1.SG par analogie) et avec la dynamique fréquentielle (nous avons mentionné dans le paragraphe précédent que la variante préconsonantique semblait s'être répandue aux contextes prépausaux puis prévocaliques par analogie). Dans un article de 1982, Morin présente encore le cas de certains [l] en français du Québec, dans la construction *ça* + VER à initiale vocalique comme dans *Ça* [l] *a pour effet…* ou *Ça* [l] *arrive souvent.*[30], qu'il analyse comme des consonnes de liaison et non comme des pronoms clitiques objets ou des déterminants (Morin, 1982 : 9, 10, 26, 27). Dans cette construction, ces [l] proviendraient analogiquement de la consonne de liaison [l] du pronom clitique sujet *elle* (mais pas de *il*), qui n'apparait que devant initiale vocalique en français de Montréal, de façon variable (Morin, 1982 : 28).

De cette façon, si l'on généralise, on peut dire que les liaisons proviennent d'anciennes consonnes finales qui ont été préservées à l'intérieur de syntagmes. Elles ont d'abord chuté devant consonne, puis à la pause, et enfin dans certains contextes prévocaliques. Nous avons cependant vu que l'évolution diachronique de la liaison n'a pas consisté qu'en une chute progressive d'anciennes consonnes finales. En effet, six dynamiques contradictoires : dynamique de syllabation ouverte, dynamique de marquage morphologique, dynamique d'enchaînement généralisé, dynamique graphique, dynamique fréquentielle et dynamique analogique ont donné lieu au phénomène de liaison tel qu'on le connaît aujourd'hui. La première a entraîné l'amuïssement puis la chute de consonnes finales, la seconde au contraire a permis le maintien de consonnes flexionnelles notamment en contexte de liaison, la troisième a encouragé la resyllabation de consonnes de liaison en attaque suite à un bouleversement prosodique majeur qui les a préservées dans ce contexte, la quatrième a créé un lien fort entre l'évolution de l'orthographe du français, avec ajout ou retrait de consonnes graphiques finales, et ce qui est aujourd'hui considéré ou non comme contexte de liaison, encourageant de ce fait la prononciation de consonnes finales graphiques ayant auparavant chuté ; avec la cinquième, les contextes prépausaux puis prévocaliques se sont calqués sur l'évolution des contextes préconsonantiques, les plus fréquents, favorisant une fois de plus la chute des consonnes qui nous intéressent et, avec la sixième, l'analogie a pu placer des liaisons dans de nouveaux contextes. Nous poursuivrons cette présentation en nous penchant sur le sort diachronique des consonnes finales latines et secondaires, avant

[30] Pour appréhender correctement ces exemples en français hexagonal, il faut retirer les [l] des exemples.

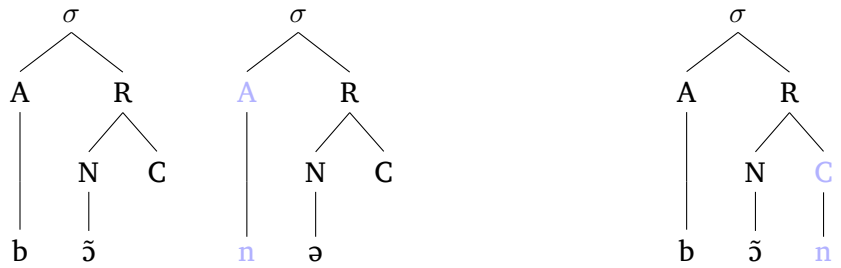
de nous intéresser à des segments vocaliques initiaux avec une destinée diachronique toute spécifique.

1.2.3.1.1.3 Sort des consonnes finales et de quelques segments initiaux spécifiques

Il est essentiel de bien comprendre que toutes les consonnes de liaison ne sont pas d'anciennes consonnes finales en latin. Elles ont pu par exemple le devenir suite à la chute d'une voyelle finale, comme pour lat. CANTĀRE > VII^{ème} s. [t͡ʃantɛr] > Français contemporain (FC) *chanter* [ʃãte] (Hambye, Année universitaire 2016-2017). Arrêtons-nous plus longuement sur le sort des diverses consonnes finales :

- Pour les huit consonnes finales du latin : [m] « s'est amuï en latin dès l'époque classique, mais se maintient dans les monosyllabes » où il se renforce en [n] avant de disparaître comme dans RĔM > X^{ème} s. [riɛ̃n] > XIII^{ème} s. [rjɛ̃n] > XVII^{ème} s. [rjɛ̃] > FC [ʁjɛ̃] (Laborderie, 1994). Il en va de même pour [n] qui ne se maintient qu'en monosyllabe (Englebert, 2009 : 126). [k] s'amuït, [r] et [l] subsistent surtout dans les monosyllabes ; seul [t] flexionnel est conservé, contrairement à [s] qui est préservé dans tous les cas (Englebert, 2009 : 126).
- Pour les consonnes finales secondaires :
  - Les consonnes [m] et [n] : le français contemporain présente des voyelles nasales : /ɑ̃/, /ɛ̃/ et /ɔ̃/ (parfois d'autres, comme /œ̃/ dans les variétés méridionales). En latin classique, il n'y a pas de voyelles nasales, mais il existe un contexte favorable à leur émergence : deux consonnes nasales, /m/ et /n/, apparaissent parfois après une voyelle orale. Le latin classique présente dix monophtongues phonologiques : /i/, /iː/, /u/, /uː/, /e/, /eː/, /o/, /oː/, /a/ et /aː/ (Laborderie, 1994). Au départ les voyelles nasales ne sont pas des phonèmes, mais par assimilation régressive de nasalité, les voyelles ou diphtongues situées devant des consonnes nasales se nasalisent entre les XI^{ème} et XIV^{ème} s. en français (Laborderie, 1994). Puis, entre le XVI^{ème} et le XVII^{ème} s., soit seule la consonne nasale est conservée, soit seule la voyelle nasale est conservée (Laborderie, 1994). Il s'agit d'un phénomène de dissimilation (Hambye, Année universitaire 2016-2017). La consonne nasale se maintient lorsqu'elle est en position forte, c'est-à-dire en initiale de syllabe (on parle de position explosive) : dans ce cas-là, la voyelle se dénasalise (elle a pu changer de timbre entre temps), comme dans BŎNA > XII^{ème} s. [bõnə] > XIII^{ème} s. [bɔ̃nə] > XVII^{ème} s. [bɔn(ə)] > FC [bɔn] *bonne* (voir figure

1.8a ; Laborderie, 1994). La voyelle nasale se maintient lorsque la consonne nasale est en position faible, c'est-à-dire en finale de syllabe (on parle de position implosive) : dans ce cas-là, après voyelles nasales, les consonnes nasales s'amuïssent, comme dans BŎNU > XII^{ème} s. [bõn] > XIII^{ème} s. [bɔ̃n] > XVII^{ème} s. [bɔ̃] *bon* (voir figure 1.8b ; Laborderie, 1994). Elles se maintiennent toutefois en liaison avec possible dénasalisation voire changement de timbre pour la voyelle précédente (voir 1.2.2.1.4), sur le modèle de la forme féminine prénominale de l'espace thématique du mot en question (voir 1.2.2.2.2).



(a) Conservation de la consonne nasale dans *bonne* (b) Conservation de la voyelle nasale dans *bon*

FIGURE 1.8 : Nasalisation en français : conservation de la voyelle nasale ou de la consonne nasale aux XVI^{ème}-XVII^{ème} s.

– La consonne [t] : le [t] final flexionnel après voyelle s'amuït entre le IX^{ème} et le XI^{ème} s. : CÁNTAT > FC *chante* (chanter.PRS.IND.3.SG), après s'être spirantisée en [θ] (Englebert, 2009 : 127) ; le [t] final flexionnel après consonne entre la fin du XII^{ème} s. et le début du XIII^{ème} s. : MŎRIT > FC *meurt* (mourir.PRS.IND.3.SG) et le [t] du participe passé, après s'être voisé au IV^{ème} s., s'être spirantisé au VI^{ème} s., être devenu final au VII^{ème} s. avant de se dévoiser, finit par s'amuïr : dans CANTĀTU, le second [t] > [d] > [ð] > [θ] > ∅ (Laborderie, 1994). Le [t] final flexionnel est uniquement maintenu en cas de liaison.
– La consonne [s] : le [s] flexionnel (verbal ou nominal) s'amuït entre la fin du XII^{ème} et le début du XIII^{ème} s. (Laborderie, 1994). Notons que les [t] devant -‹s› ont la particularité de former au préalable l'affriquée [t͡s], par la suite réduite à [s] (réduction des affriquées au XIII^{ème} s. ; Laborderie, 1994), qui s'amuïra de la même façon (Morin, 1986 : 168). Le [s] flexionnel (verbal ou nominal) n'est maintenu qu'en cas de liaison dans sa variante voisée [z].
– Les consonnes voisées ayant une variante non voisée se dévoisent (Englebert,

2009 : 126-127). Ainsi, [b], [d], [g], [ð] (résultat d'une spirantisation), [v] et [z] deviennent aux VII^{ème}-VIII^{ème} s. : [p], [t], [k], [θ], [f] et [s] (Laborderie, 1994). La prononciation dévoisée des occlusives survit en liaison (Hornsby, 2020 : 47).

– La consonne [r] : elle s'amuït seulement au XVI^{ème} s., mais est rétablie par les grammairiens au XVII^{ème} s. dans la plupart des cas, comme dans les infinitifs en *-ir* et *-oir* (mais pas ceux en *-er*), où elle est marque flexionnelle, ou encore dans les mots en *-eur* (Laborderie, 1994). Son rétablissement à l'oral pour les infinitifs semble s'être fait « par analogie avec des verbes dont le [r] [s'était] maintenu, n'étant pas final (*écrire*, *dire* ; *boire*, *croire*) » (Englebert, 2009 : 131), puisque suivi d'un schwa qui se labialise en [œ] au XV^{ème} s. avant de s'amuïr au XVII^{ème} s. Ainsi cette consonne ne peut être liaisonnante que pour les infinitifs en *-er*. Rappelons au passage qu'en latin, les ‹r› étaient des trilles alvéolaires qui se sont d'abord postériorisées en trilles uvulaires avant de s'affaiblir en fricatives uvulaires en français standard, tout ceci entre le XIII^{ème} et le XVII^{ème} s. : /r/ > /ʀ/ > /ʁ/ (Hambye, Année universitaire 2016-2017).

Nous venons de voir que ces anciennes consonnes finales ont évolué différemment : si certaines ont complètement disparu (liaisons non attestées en corpus, comme l'exemple souvent cité de *et* + M2), d'autres ont atteint un état où elles ne sont visibles qu'en contexte de liaison, quand d'autres peuvent aussi avoir été conservées à la pause et devant initiale consonantique sous diverses formes : le fonctionnement des numéraux *six* et *dix*, avec trois formes : une en isolation, une devant initiale vocalique et une devant initiale consonantique en est un bon exemple (voir Type 4 du tableau 2.2). Selon Marchello-Nizia et al. (2020 : 484), « [c]e régime ternaire n'a pas perduré en français moderne. [...] Dans la période contemporaine, on note la réapparition variable d'un régime ternaire pour *six* et plus récemment pour *dix* ». On est donc face à un changement diachronique qui n'est pas uniforme.

Prenons aussi le temps de nous attarder sur quelques contextes particuliers en initiale de M2 :

- Commençons par les ‹h›- disjonctifs, que nous avions mentionnés en 1.2.2.1.4. Les mots à ‹h› graphiques initiaux peuvent être divisés en deux groupes : ceux avec lesquels la liaison et l'élision sont habituellement faites (*habitude*, *homme*, *honneur*, *horloge*, *humeur*...), et ceux avec lesquels la liaison et l'élision ne sont habituellement pas réalisées (*hache*, *haïr*, *héron*, *hêtre*, *honte*, *housse*, *houx*...). La

première série regroupe des mots d'origine latine, tandis que la seconde regroupe des mots d'origine germanique (Laborderie, 1994). Au niveau graphique, l'ancien français ne notait pas les ‹h›- d'origine latine mais bien ceux d'origine germanique (Marchello-Nizia et al., 2020 : 347). Dès les XIV^{ème} et XV^{ème} s., la plupart des ‹h›- latins ont été rétablis à l'écrit, hormis quelques exceptions comme *avoir* < lat. HABERE (Marchello-Nizia et al., 2020 : 348 ; CNRTL, 2012a). Dans le latin introduit en Gaule, les ‹h›- n'étaient déjà plus prononcés, puisqu'ils s'étaient amuïs dès le I^{er} s (Laborderie, 1994). Certains ‹h›- sur des mots d'origine latine ne sont en outre pas étymologiques (par exemple, *huile* < lat. OLEUM ; Thibault, Année universitaire 2019-2020), ayant été ajoutés à une époque ou ‹v› pouvait autant correspondre à la consonne /v/, aux voyelles /u/ ou /y/ ou aux glissantes /w/ ou /ɥ/. Sa présence devant ‹v› indiquait une prononciation non consonantique (*huile* 'huile' vs *vile* 'ville' ; Thibault, Année universitaire 2019-2020). L'usage systématique du graphème ‹v› pour la prononciation [v] ne s'est établi qu'au XVII^{ème} s (Laborderie, 1994). Par contre, les ‹h›- des emprunts germaniques, arrivés plus tardivement, au V^{ème} s. (Englebert, 2009), avaient encore une réalité phonétique : on parle à tort d'aspiration mais il s'agissait plutôt d'une fricative glottale non voisée [h] qui a pu être conservée chez certains locuteurs francophones jusqu'à nos jours (elle s'est parfois transformée en occlusive glottale [ʔ] ; Thibault, Année universitaire 2019-2020). L'amuïssement de cette consonne s'est produit dans la variété standard dès le début du XIII^{ème} s. (des graphies sans ‹h›- apparaissent) pour s'achever au XVII^{ème} s. (Marchello-Nizia et al., 2020 : 347-348). Par conséquent, pour la deuxième série mentionnée plus haut, on peut parler d'une influence du superstrat germanique sur la phonologie du français contemporain, qui ne se manifeste plus que par une disjonction lors d'une liaison ou d'une élision. Il faut toutefois bien noter que tous les ‹h›- disjonctifs ne sont pas d'origine germanique. De rares cas, que je ne mentionnerai pas ici, proviennent du latin (voir Marchello-Nizia et al., 2020 : 348). Un phénomène similaire s'est produit pour des emprunts arrivés plus récemment, provenant d'autres familles de langue comme avec l'arabe (*houmous*, *harem* via le turc…) ou à nouveau de la sous-famille germanique comme avec l'anglais (*hamburger*, *handicap*…). Quelle que soit l'origine des mots à ‹h›- disjonctifs, leur prononciation peut déjà avoir évolué dans les pratiques (comme pour les mots *hamburger* ou *handicap* précités). On parle parfois de ***psilose*** quand une liaison est produite dans un contexte historiquement disjonctif, ce qui semble peu approprié, puisque le phénomène de psilose correspond à la « [d]isparition de l'aspiration à l'initiale d'un mot » (CNRTL, 2012l), et nous venons de voir que

les ‹h› disjonctifs étaient improprement qualifiés d'*aspirés*. Enfin, la disjonction ne concerne pas que des mots à ‹h› graphiques initiaux, comme nous l'avions vu en 1.2.2.1.4, mais aussi des mots à glissantes initiales (*whisky* qui vient de l'irlandais via l'anglais, *yéti* du tibétain…), des noms de lettres ou sigles, des nombres cardinaux ou ordinaux, des entités nommées ou encore des interjections.

- Outre les mots à ‹h› disjonctifs initiaux, on trouve par exemple les **prothèses vocaliques** (ou prosthèses vocaliques) parmi les autres voyelles initiales non étymologiques en français. Dès le Ier s., un ĭ apparaît devant les groupes consonantiques complexes /s/ + occlusive dans les mots d'origine populaire, qui évoluera vers [e] par la suite (comme pour STA̍BULA > [ĭs'tabula] > [es'tabula] > Ancien français (AF) *estable* > *étable* (Englebert, 2009 ; Laborderie, 1994). La liaison touchera ces mots.
- D'autres mots à initiale vocalique non étymologique proviennent encore d'**agglutinations**, ou réunions d'éléments phonétiques appartenant à des morphèmes différents en un seul élément morphologique, comme dans l'exemple 25. Les agglutinations sont un cas particulier de **métanalyses**, ou resegmentations de mots. Ici, on peut supposer qu'*aperdrix* (lat. PERDICEM, Accusatif (ACC) de PERDIX ; CNRTL, 2012j) résulte d'une métanalyse de *la perdrix* en *l'aperdrix*. Cela peut se produire enfin avec le phénomène inverse, un autre cas de métanalyse, appelé **déglutination**, ou dissociation d'éléments phonétiques appartenant à un seul morphème en plusieurs éléments morphologiques. Ces différents types de métanalyses ont pu se lexicaliser, comme pour les agglutinations *lierre* (AF *iere* ; CNRTL, 2012e), *luette* (AF *uete*, cf emprunts savants *uvule* et *uvulaire* ; CNRTL, 2012f), *nombril* (AF *umblil*, cf emprunt savant *ombilical* ; CNRTL, 2012g) ou encore *tante* (AF *ante* ; CNRTL, 2012m ; cf aussi *aunt* en anglais, *t'* étant l'ancienne forme prévocalique élidée du déterminant féminin[31]) ou les déglutinations *griotte* (provençal *agriota* ; CNRTL, 2012d) ou *orange* (arabe *nārang(a)* ; CNRTL, 2012h)[32]. Certaines agglutinations (comme *aperdrix*) ou déglutinations (comme *orange*) sont liaisonnantes bien que ces mots ne soient pas étymologiquement à voyelle initiale. Au contraire, des agglutinations (comme *lierre*, *luette*, *nombril*, *tante*) ou déglutinations (comme *griotte*) ne sont pas liaisonnantes alors que ces mots étaient étymologiquement à initiale vocalique.

(25) *SGR : tuer des lièvr(e)s des aperdrix.

[31] Voir 1.2.2.2.4 où l'évolution des déterminants unipossessifs féminins prévocaliques est retracée.

[32] Pour plus d'exemples, consulter Reece, 2009.

(JD_a062_07, 43810, Virginie Granouillet)

En résumé, nous avons examiné le sort de nombreuses consonnes finales, latines ou secondaires, qui peuvent avoir donné lieu à des liaisons en français. Le caractère flexionnel de certaines consonnes les a préservées plus longtemps. Nous avons vu que l'apparition de voyelles nasales en français avait amené à la chute de certaines consonnes nasales finales, maintenues en liaison, que des occlusives et fricatives ayant une variante non voisée s'étaient dévoisées en fin de mot, ce qu'on retrouve pour les occlusives en liaison aujourd'hui et après avoir rappelé l'évolution diachronique du phonème /ʁ/ en français, nous avons vu que des [r] finaux avaient été rétablis par des grammairiens. Les changements ont eu lieu à des étapes différentes et ont abouti à diverses situations : disparition totale, maintien en liaison ou mise en place d'un régime ternaire (sans compter les finales toujours lexicales). Pour finir, nous avons retracé l'origine des différents mots à ‹h› graphiques initiaux afin de comprendre pourquoi certains avaient un comportement disjonctif et mentionné l'existence des prothèses vocaliques, de phénomènes d'agglutination et de déglutination qui ont modifié l'initiale de certains mots, ceci ayant des conséquences sur le caractère liaisonnant ou non des mots en question. Maintenant que nous connaissons l'histoire générale du phénomène de liaison, nous allons nous intéresser au sort des liaisons en microdiachronie. Pour observer des évolutions plus récentes, nous disposons à l'heure actuelle de corpus oraux à interroger.

1.2.3.1.1.4 La liaison en français contemporain : corpus oraux microdiachroniques

L'idée reçue selon laquelle les liaisons seraient de moins en moins réalisées est assez répandue (Bauche, 1920 : 56 ; Grammont, 1933 : 130 ; Chigarevskaïa, 1973 : 162 ; Müller, 1985 : 229 ; Tranel, 1987 : 169 ; Léon, 1992 : 246 ; Thomas, 1998 : 545-546). Certains auteurs restreignent cette affirmation aux liaisons variables (Desrochers, 1994 : 280). En effet, comme le notait Encrevé (1983 : 49), « [l]es liaisons obligatoires ne créant aucune distinction, ni positive ni négative, ne sont pas socialement repérées comme des liaisons, au point que bien des auteurs parlent depuis longtemps de dépérissement de la liaison [...] ». Qu'en est-il réellement dans les corpus ? Constate-t-on une baisse générale, uniquement pour les facultatives, ou cette idée reçue de diminution n'est pas confirmée ?

Avant de présenter les résultats de diverses études, je ferai mention d'un aspect méthodologique essentiel : dans une perspective labovienne du changement linguistique, il

faut distinguer les **études en temps réel**, lorsqu'on observe la communauté linguistique à deux moments temporels distincts (Labov, 1994 : 73), des **études en temps apparent**, lorsqu'on compare les productions langagières des différentes générations de la communauté linguistique (Labov, 1994 : 28). Si la notion de génération se rapporte communément à la classe d'âge à laquelle chaque individu appartient, selon Encrevé (1983 : 60) :

> [...] la notion de génération [doit être] entendue [...] non comme classe d'âge biologique mais historique. Une génération réunit des agents ayant participé (différemment) à une même histoire sociale : s'agissant de l'acquisition de la langue, une même génération de locuteurs a été exposée à un même état du système d'enseignement (défini par le mode de recrutement et de formation des enseignants, par le mode d'accès des enseignés à tel niveau d'enseignement, par la durée de la scolarité obligatoire, etc.

Pour les sociolinguistes, le changement linguistique ne peut être réduit à la dimension diachronique, car il y a bien souvent, comme nous allons le voir, intersection avec d'autres dimensions.

En ce qui concerne la parole publique en diachronie, Encrevé (1983, 1988) s'est intéressé aux liaisons non enchaînées et voulait entre autres savoir si ce phénomène, qu'il s'était mis à noter mais qu'aucun de ses contemporains n'avait étudié de façon systématique et approfondie, était nouveau et en augmentation (Encrevé, 1983 : 39). De la sorte, ce chercheur produit la première étude microdiachronique du phénomène de liaison. Travaillant sur un corpus de discours publics de personnalités politiques françaises, il note que chez les chefs d'État de 1928 à 1981, la norme serait passée de la réalisation quasi-systématique de l'enchaînement à la réalisation variable de celui-ci (Encrevé, 1983 : 49, 53-54). Il remarque aussi qu'entre 1978 et 1981, le non-enchaînement aurait été généralement en hausse (7/11) ou constant (3/11) chez les personnalités politiques qu'il étudie (Encrevé, 1983 : 53).

Laks (2008, 2014) et Laks et Peuvergne (2017), sur la lancée d'Encrevé, s'intéressent à leur tour à la liaison dans la parole publique des personnalités politiques françaises avec le corpus HPOL, qui comprend deux volets : un premier couvrant la période 1908-1998, HPOL1, et un second portant de 1999 à 2015, HPOL2 (Laks & Peuvergne, 2017 : 57). Au cours du XX^{ème} siècle, le taux de réalisation des liaisons serait resté stable dans HPOL1, ce qui masque pourtant une évolution plus importante : les liaisons variables ont d'abord chuté jusque dans les années 1970, avant de remonter brusquement, changement qui coïnciderait avec le moment où le non-enchaînement est devenu significatif

(Laks, 2008 : 245). Ce dernier atteint son pic en 1981 (12,6% des facultatives), avant de régresser et de se marginaliser (3,7% en moyenne dans HPOL2, c'est-à-dire entre 1999 et 2015 ; Laks & Peuvergne, 2017 : 63). Grâce aux annotations mises en place dans HPOL1, Laks peut étudier l'évolution diachronique du caractère liaisonnant des M1 en fonction par exemple de leur nombre de syllabes ou de leur nombre : il estime que les polysyllabes (deux à quatre syllabes) seraient devenus moins liaisonnants et les monosyllabes plus liaisonnants au fil du temps, tout comme les mots singuliers seraient devenus moins liaisonnants et les mots pluriels plus liaisonnants au fil du temps (Laks, 2008 : 256). Cela me semble toutefois discutable car les chiffres fluctuent beaucoup, ce qui nous impose de vérifier la tendance générale et si différence il y a, que cette dernière est significative. Dans HPOL2, Laks et Peuvergne (2017 : 64) présentent une étude en temps apparent, en distinguant les locuteurs nés entre 1927 et 1945 (avant-guerre), de ceux nés entre 1946 et 1958 (IV^{ème} République) et de ceux nés entre 1961 et 1976 (V^{ème} République)[33]. Le taux de réalisation des liaisons variables entre la première (267/680 ; 39%) et la deuxième génération (673/2044 ; 33%) a baissé de façon significative et il en va de même entre la seconde et la dernière génération (373/1353 ; 28%), tout particulièrement dans les discours, où la différence de réalisation globale est significative (génération 1 : 319/489 ; 65% ; génération 3 : 490/939 ; 52% ; Laks & Peuvergne, 2017 : 66), la dimension diachronique croisant ici la dimension diaphasique. Le changement est mené par les femmes qui ont un taux de réalisation global inférieur aux hommes (F 1367/2578 ; 53% vs H 2312/3909 ; 59% : Laks & Peuvergne, 2017 : 66), la dimension diachronique croisant cette fois-ci la dimension diagénique.

Toujours en ce qui concerne la parole publique, Smith (1996), cité par Armstrong et Pooley (2010 : 114), a effectué une comparaison de parole journalistique à trente ans d'intervalle (années 1960 vs années 1990) en comparant les données d'Ågren (1973) à un corpus qu'il a recueilli. Il aurait noté une diminution des liaisons variables (de 62% à 47%), notamment significative après les formes du verbe *être* conjugué (de 87,9% (N = 3858) à 78,8% (N = 1301)), après les formes du verbe *avoir* conjugué (de 51,4% (N = 1140) à 28% (N = 411)), après les formes des « semi-auxiliaires » conjugués (*falloir*, *devoir*, *pouvoir* : de 50,2% (N = 851) à 25,8% (N = 407)) et après les adverbes et prépositions polysyllabiques (*pendant*, *toujours* : de 40,3% (N = 988) à 12,8% (N = 234)).

Hutin et al. (2022) ont elles aussi repris les données d'Ågren (1973), mises en perspective cette fois-ci avec les données du corpus Évaluation en Traitement Automatique de la ParolE (ETAPE)[34]. Elles observent, à cinquante ans d'intervalle (années 1960 vs

[33] Les délimitations des deux derniers groupes varient légèrement dans Laks, 2014.

[34] ETAPE est un corpus de parole radiophonique et télévisuelle en français (Gravier et al., 2012).

années 2010), une diminution significative des liaisons variables réalisées (de 96% (1388/1488) à 68% (511/757)) après les formes du verbe *être* conjuguées (*suis*, *est*, *sommes*, *êtes*, *sont*, *était*, *fut* et *serait*) + *un/une*.

Ashby (2003) a été le premier à s'intéresser à la liaison en diachronie dans la parole privée, à l'aide d'un corpus de locuteurs tourangeaux (de la ville de Tours) équilibré hommes/femmes et permettant l'opposition de classes socio-économiques (basse vs haute) et de classes d'âge (locuteurs jeunes : 14-22 ans vs locuteurs âgés : 51-64 ans). Ce corpus comprend deux volets, l'un enregistré en 1976 (16 locuteurs) et l'autre en 1995 (22 locuteurs ; Ashby, 2003). Le linguiste fait le choix de s'intéresser directement à la liaison variable et observe à vingt ans d'intervalle une diminution du taux de réalisation de 36% (586/1617) à 28% (290/1034), changement linguistique qui semble se produire dans certaines constructions spécifiques, notamment *est* + M2 (de 58% (304/522) de réalisations à 37% (150/400)) et *pas* + M2 (de 12% (14/112) de réalisations à 4% (5/120) ; Ashby, 2003). Le changement est mené par les hommes (de 37% (281/741) de liaisons variables réalisées à 21% (92/431), taux statistiquement significatif par rapport aux réalisations des femmes en 1995 qui s'élèvent à 33% (198/603) ; Ashby, 2003). Seules les productions de la couche sociale haute évoluent (de 44% (467/1047) de liaisons variables réalisées à 33% (207/616)), celles de la couche sociale basse étant stabilisées autour de 20% (Ashby, 2003). Enfin, les productions des deux classes d'âge baissent de façon significative pendant la période (classe d'âge supérieure : de 47% (388/813) à 38% (245/635) ; classe d'âge inférieure : de 24% (198/804) à 11% (45/399)), mais les jeunes ont des taux de réalisation plus bas que les locuteurs âgés (Ashby, 2003).

Laks (2008 : 240-241, 243) a également comparé BRU, une enquête menée en 1912 et 1913 par Ferdinand Brunot auprès de trois ouvriers nés en 1848, 1873 et 1875, VIL, une enquête qu'il a lui-même effectuée à Villejuif en 1975 auprès de huit adolescents issus de classes populaires[35] et le corpus PFC (enregistré entre 1996 et 2015[36]). Si les taux de réalisation des liaisons globaux de PFC et de BRU se sont révélés similaires, contrairement à l'affirmation de Laks (2008 : 243), VIL se comporte différemment. Par ailleurs, si les taux de réalisation de PFC et VIL diffèrent bien de celui d'HPOL1, ce dernier ne se démarque pas de celui de BRU, à nouveau, à l'encontre de ce qu'avance Laks[37] (2008 : 243). Malgré quelques conclusions hâtives, cet auteur nous met en garde : la

[35]Le corpus est abrégé VILL dans Laks, 2014 puis renommmé LAKS dans Laks & Peuvergne, 2017.

[36]Selon les informations disponibles en juin 2018 sur le site https ://research.projet-pfc.net/, sachant que l'article date de 2008.

[37]Il se pourrait que la confrontation d'un corpus synchronique à un corpus diachronique sur un siècle influence ce résultat. Il aurait pu être préférable de comparer BRU aux données d'HPOL1 de la même époque.

différence de réalisation au niveau des liaisons entre les corpus ne relèverait pas d'un changement diachronique, mais de variation diastratique et diaphasique (Laks, 2008 : 245), puisqu'un taux élevé de liaisons facultatives, propre à un technolecte (la façon de parler des personnalités politiques) et à un style (le discours politique) s'opposerait à des pratiques moins liaisonnantes, propres à la population générale, dans un style spécifique : la conversation. On pourrait ajouter à cela le rôle sans équivoque de la dimension diamésique, les discours politiques ayant un fort lien avec l'écrit. La parole politique tend pourtant à se rapprocher de la parole ordinaire en diachronie, puisque les données de Laks et Peuvergne (2017 : 61, 66), lors de leur étude en temps apparent opposant trois générations de personnalités politiques, montrent que la dernière présente un taux de réalisation des liaisons global (1104/2100 ; 53%) comparable aux productions non professionnelles (25534/53594 ; 47,7% dans le corpus PFC), alors que la différence était significative pour les deux générations précédentes, convergence qui pourrait être due au prestige moindre de l'écrit pour des locuteur nés après 1968 (Laks & Peuvergne, 2017 : 65).

Enfin, Dugua et Baude (2017), s'intéressent eux aussi à la liaison dans la parole privée de locuteurs orléanais avec le corpus des ESLO, qui comprend deux volets : un premier enregistré dans les années 1970 (1968-1971), ESLO1, et un second dans les années 2010 (2008-2012), ESLO2 (Eshkol-Taravella et al., 2011 : 19, 48). Le premier volet a été constitué « par une équipe [...] franco-britannique [...] [avec] pour objectif de fournir un corpus représentatif pour le français tel qu'il est parlé, dans le but premier de constituer une méthode de langue fondée sur des documents authentiques » (Baude & Dugua, 2015 : 355). À l'aide du module Diachronie[38], les productions linguistiques de sept mêmes locuteurs peuvent être comparées à quarante ans d'intervalle (vingt minutes ont été retenues pour chaque enregistrement ; Dugua & Baude, 2017 : 46). Si le taux global de réalisation des liaisons, celui des liaisons invariables mais aussi celui des liaisons variables seraient restés stables entre les deux périodes, une variation inter-individuelle moins étendue lors de la réalisation des liaisons variables dans ESLO2 comparé à ESLO1 « pourrait être interprétée comme une plus grande convergence des usages » actuelle (Dugua & Baude, 2017 : 48, 51). Par ailleurs, la figure 1.9 laisse penser que ces chiffres globaux, tout particulièrement ici le taux de réalisation des liaisons variables, pourraient occulter des évolutions individuelles entre les deux périodes : augmentation, stabilité ou diminution selon les locuteurs (même s'il faudrait vérifier que chaque augmentation ou diminution est statistiquement significative ; Dugua &

[38] Ou ESLO-DIA(chronie) dans Baude & Dugua, 2015 : 356.

Baude, 2017 : 48-49). Les auteurs remarquent enfin que les locuteurs dont l'usage de la liaison variable a augmenté appartiennent à des catégories socio-économiques basses alors que ceux pour lesquels elle est restée stable appartiennent à des catégories hautes, quand ceux pour lesquels elle a baissé appartiennent à diverses catégories, ce qui nous rappelle que les différents types de variation s'entremêlent, le changement diachronique dont il est question ici relevant également de la variation diastratique (Dugua & Baude, 2017 : 52).

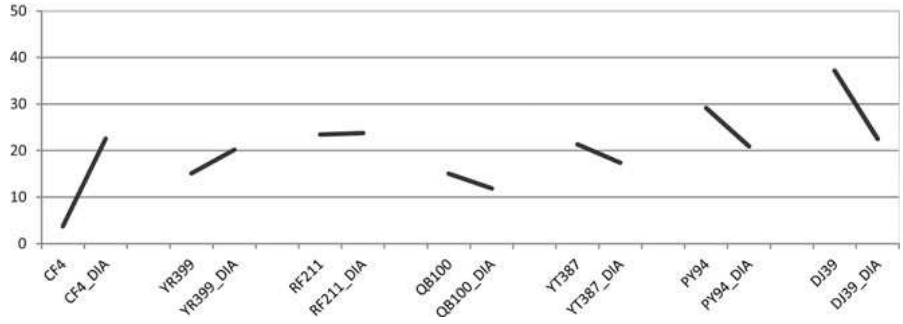


FIGURE 1.9 : Évolution des taux de liaisons variables réalisées entre les deux périodes d'enregistrement chez les sept locuteurs du module Diachronie dans le corpus des ESLO (Dugua et al., 2017 : 48)

En guise de conclusion, contrairement à une idée reçue impressionniste selon laquelle les liaisons seraient de moins en moins réalisées, le taux global de réalisation aurait peu évolué en microdiachronie, que ce soit chez les personnalités politiques en temps réel au XX^{ème} s. (HPOL1 ; Laks, 2008) ou chez les locuteurs orléanais en temps réel depuis les années 1970 (Dugua & Baude, 2017). Une analyse plus fine exhibe toutefois des tendances baissière comme haussière dans la réalisation des liaisons variables chez les personnalités politiques durant le siècle dernier (HPOL1 ; Laks, 2008), d'où la nécessité d'étudier une fenêtre temporelle large. Par exemple, Encrevé (1983) a publié son travail sur le non-enchaînement en temps réel chez les personnalités politiques dans les années 1980, à une époque où le phénomène était à son apogée, avant qu'une étude de Laks et Peuvergne en temps réel (2017) ne montre qu'il s'est marginalisé depuis lors. Les résultats d'Ashby (2003), présentant une diminution des liaisons variables en temps réel dans la parole privée entre 1976 et 1995 ou ceux de Smith (1996, cf Armstrong & Pooley, 2010), présentant une diminution des liaisons variables en temps réel dans la parole journalistique entre les années 1960 et 1990, sont de la même façon à prendre avec précaution. Une analyse fine révèle par ailleurs, avec l'étude en temps apparent

de Laks et Peuvergne (2017), qu'une baisse significative du taux de liaisons variables dans la parole politique s'est produite à chaque génération, menée par les femmes, tout particulièrement dans les discours, quand dans l'étude en temps réel d'Ashby (2003), le changement est mené par les hommes, alors que la classe sociale inférieure et les jeunes (temps apparent) sont les moins liaisonnants. Tout ceci rappelle à juste titre que les dimensions de la variation interagissent. Des taux comparables entre les ouvriers de BRU au début du siècle et les locuteurs de PFC à la fin du siècle et une opposition entre les productions des locuteurs de VIL et PFC et celles des personnalités politiques d'HPOL1 (Laks, 2008) pourraient à nouveau souligner la prédominance d'autres types de variation sur la dimension diachronique à proprement parler, ce que corrobore l'étude en temps apparent de Laks et Peuvergne (2017) où les productions de la dernière génération de personnalités politiques sont désormais comparables à celles du corpus variationniste PFC. Pour terminer, une analyse fine montre que le changement linguistique s'observe dans des constructions spécifiques (Smith, 1996, cité dans Armstrong & Pooley, 2010 ; Ashby, 2003 ; Hutin et al., 2022 en temps réel), que la variation interindividuelle pourrait être moindre aujourd'hui et qu'au cours de la vie, des évolutions individuelles à la hausse comme à la baisse seraient possibles (Dugua et al., 2017). Nous continuerons à nous intéresser à la liaison en diachronie, mais pour lors sous l'angle de la morphologie.

1.2.3.1.2 Variation diachronique, liaison et morphologie

1.2.3.1.2.1 Statut morphologique des consonnes de liaison en diachronie

Historiquement, exceptées les liaisons par analogie, les consonnes de liaison proviennent du mot liaisonnant. Nous avions vu en 1.2.2.2.1 que le phénomène de liaison pouvait être abordé comme un cas d'alternance entre radical long et radical court. Dans la diachronie du français, le radical court constitue une innovation, qui finit par devenir le radical par défaut (Morin, 1986 : 197). Nous avions vu en 1.2.3.1.1.2 que diachroniquement, celui-ci s'est développé d'abord devant initiale consonantique puis à la pause (il est actuellement systématique dans ces contextes), avant de s'étendre à certains contextes prévocaliques (en cas de non-liaison ; Marchello-Nizia et al., 2020 : 399). Le radical long au contraire ne se maintient qu'en tant que morphème lié (Morin, 1986 : 197). Il existe pourtant une exception que j'avais mentionnée en 1.2.2.2.4 : la forme élidée des déterminants unipossessifs féminins (du type *m'amie*) en ancien français a été remplacée par la forme prévocalique masculine (du type *mon amie*) en français

moderne (Chevalier & Fréchette, 2004 : 7) : ici, le radical long constitue donc une innovation, générant un contexte de liaison là où il n'y en avait pas auparavant.

D'autre part, nous avions vu que les consonnes de liaisons pouvaient être considérées comme des morphèmes flexionnels. De la sorte, il y a tout d'abord une réanalyse de consonnes finales de mots liaisonnants comme des suffixes. Pour Morin (1986 : 195), ceux-ci peuvent ensuite être réanalysés comme des préfixes, ce que lui semble illustrer l'exemple suivant : *c'est quoi comme* [z] *arbres ?* (en sachant que *comme* ne présente habituellement pas de marque de nombre). Nous serions alors de nouveau face à une innovation. Pour s'en convaincre, nous verrons dans le point suivant (voir 1.2.3.1.2.2) que la marque de pluriel [z] en français provient d'un suffixe verbal flexionnel -[s] de l'ancien français, amalgame marquant à la fois le cas et le nombre (OBL.PL), avec donc déjà une valeur plurielle, -[s] qu'on relève cependant ailleurs dans le paradigme sans valeur plurielle (NOM.SG). Celui-ci est issu du suffixe latin *-os*, amalgame marquant à nouveau le cas et le nombre (ACC.PL). La morphémisation reste cependant incomplète car elle n'est pas ajoutée systématiquement à tous les mots pluriels et n'apparaît pas hors liaison. Nous verrons de même pour les marques de personne (en 1.2.3.1.2.2) qu'en latin, -[s] était présent en finale des 2.SG, 1.PL et 2.PL. -‹s› graphique s'est ensuite étendu à certaines 1.SG (à l'imparfait ou encore au présent de l'indicatif et au passé simple, mais cette fois-ci uniquement pour les verbes autres que ceux en -‹er›/-‹ier›, généralement depuis la 2.SG) et 2.SG (aux formes impératives) où il peut aujourd'hui être prononcé [z] en liaison, marque apparaissant ainsi après des verbes conjugués aux premières et deuxièmes personne. Au contraire, -[t], qui était présent en finale des 3.SG et 3.PL, continue d'apparaître aujourd'hui en liaison après des troisièmes personnes. La morphémisation reste cependant incomplète car elle n'est pas ajoutée systématiquement à tous les verbes conjugués et n'apparaît pas hors liaison.

Historiquement, les radicaux courts ont donc progressivement remplacé les radicaux longs, maintenus uniquement en liaison (à l'exception du cas des déterminants unipossessifs féminins), et la marque de pluriel [z] et les marques de personne [z] et [t] ont été morphémisées, bien que leur utilisation ne soit pas systématique. Comme annoncé, c'est l'origine de ces marques de pluriel et de personne qui nous intéressera à présent.

1.2.3.1.2.2 Origine des consonnes flexionnelles

Nous nous arrêterons pour commencer sur un cas de flexion nominale, avant de nous arrêter sur les cas de flexion verbale.

En ce qui concerne la flexion nominale, l'affaiblissement des syllabes finales mentionné en 1.2.3.1.1.2 a aussi eu pour conséquence, comme nous le voyons dans le tableau 1.2, de faire passer le système casuel de six cas en latin[39] à deux cas en ancien français. Dans ce chronolecte, le -‹s› correspond ainsi à deux marques casuelles nominales pour le masculin. Il est intéressant de savoir que c'est le -‹s› du cas régime (ou oblique) pluriel de l'ancien français, issu de l'accusatif pluriel du latin, qui deviendra la marque de pluriel régulier en français (Perret, 1998).

	latin		ancien français	
SG	NOM	*murus*	NOM	**murs**
	ACC	*murum*	OBL	**mur**
	GEN	*muri*		
	DAT/ABL	*muro*		
PL	NOM	*muri*	NOM	**mur**
	ACC	*muros*	OBL	**murs**
	GEN	*murorum*		
	DAT/ABL	*muris*		

TABLEAU 1.2 : *Murus* : from Latin to Old French (Hornsby, 2020 : 48)

Pour ce qui est de la flexion verbale, par défaut, l'accent sera mis sur le présent de l'indicatif, et quand il sera question d'un autre temps ou mode dans ce qui suivra, cela sera indiqué. /z/ n'est pas étymologique après les 1.SG : en latin, la marque de 1.SG était -[o], marque qui a disparu en ancien français, et celle de l'imparfait était -[m] (Marchello-Nizia et al., 2020 : 750, 758). Par la suite, -[o] tend à être remplacée par -‹e› (pour les verbes en -‹er› ou -‹ier›)[40] ou -‹s› (pour les verbes autres que ceux en -‹er› ou -‹ier›) qui se développe à partir du XIV^{ème} s. par analogie avec des verbes inchoatifs comme fr. 1.SG *finis* (lat. 1.SG FINIO) où ‹s›- appartient à la base mais n'y est pas étymologique[41], par analogie avec des verbes du troisième groupe comme fr. 1.SG *puis* (lat. 1.SG *POSSYO) où ‹s›- appartient à la base et y est étymologique ou par analogie avec

[39] Le vocatif est absent du tableau 1.2.

[40] Par analogie avec les verbes ayant développé un -[ə] de soutien (après un groupe consonantique occlusive + liquide, comme pour lat. 1.SG TREMULO > fr. 1.SG *tremble*, ou après la séquence consonne + [d͡ʒ], comme pour lat. 1.SG SOMNIO > fr. 1.SG *songe*) ou encore par analogie avec les 2.SG et 3.SG des verbes étymologiquement en -‹are› qui ont maintenu un -[ə] issu de la voyelle désinentielle [a] qui s'était affaiblie au VII^{ème} s. à ces personnes comme lat. 2.SG AMAS & 3.SG AMAT > fr. 2.SG *aimes* & 3.SG *aime*, d'abord pour les formes à radical consonantique comme 1.SG *aime* (changement achevé au début du XV^{ème} s.), puis pour ceux à radical vocalique comme 1.SG *prie* (changement achevé au XVI^{ème} s. ; Marchello-Nizia et al., 2020 : 749-750).

[41] « Les verbes inchoatifs [...] présentent une base élargie en *-ss-*, issu du suffixe latin *-sc-* qui avait été étendu à de nombreux verbes, à l'origine de la conjugaison dite du 2^{e} groupe [...] » (Marchello-Nizia et al., 2020 : 754).

la 2.SG, par opposition à la 3.SG : Andrieux et Baumgartner (citées par Morin & Bonin, 1992 : 36) proposent que : « l'extension du [-s] correspondrait à une tentative d'opposer morphologiquement les personnes d'allocution (1sg/2sg) à la 3sg ». Quelque soit l'origine du -‹s›, celui-ci apparaît d'abord pour les formes à radical consonantique comme 1.SG *viens*, puis pour les formes à radical vocalique comme 1.SG *crois* (Marchello-Nizia et al., 2020 : 751). Comme le rappellent Marchello-Nizia et al. (2020 : 780), -[s] et -[t] flexionnels ne se maintiennent que jusqu'au XIII^{ème} s. Les chronolectes postérieurs à l'ancien français ne présentent plus de correspondance graphie/phonie stricte, et les enregistrements les plus anciens remontent seulement à un peu plus d'un siècle. Ainsi, on fera attention à ne pas confondre une innovation graphique, le -‹s› qui s'étend par analogie aux 1.SG, avec une évolution phonétique : on ne peut certifier que les liaisons étaient systématiquement produites après des verbes autres que ceux en -‹er› ou -‹ier›. Les réalisations qu'on observe aujourd'hui pourraient relever d'un orthographisme non systématique s'étant produit à une époque spécifique.
Au passé simple, un -‹s› se développe aussi à la 1.SG sur les verbes autres que ceux en -‹er›/-‹ier›.
À l'imparfait, l'ajout d'un -‹s› à la 1.SG, sur le modèle des 2.SG, s'impose peu à peu (concurrence entre -‹oie›, désinence majoritaire[42] au XIV^{ème} s., -‹oye›, désinence majoritaire[42] au XV^{ème} s., -‹oi›, -‹oy›, -‹ois›, désinence majoritaire[42] au XVI^{ème} s., -‹oys› et -‹ais›, qui est devenue la désinence du français contemporain ; Marchello-Nizia et al., 2020 : 785-787).
Pour ce qui est du futur et du conditionnel présent, Marchello-Nizia et al. (2020 : 789) rappellent qu'ils sont de formation romane, étant issus :

> d[e] formations périphrastiques associant l'infinitif atone du verbe et un auxiliaire accentué et conjugué, en général *habere*. L'infinitif s'est progressivement soudé aux formes réduites de l'auxiliaire, conjugué au présent de l'indicatif pour le futur et à l'imparfait de l'indicatif pour le conditionnel.

L'évolution des marques verbales du présent et de l'imparfait de l'indicatif suffisent ainsi à expliquer celles du futur de l'indicatif et du conditionnel présent.

En ce qui concerne la 2.SG, le -[s] est présent dans les formes latines, puis également en ancien français où il est parfois précédé d'un [ə] de soutien pour les verbes en -‹er› ou

[42] Marchello-Nizia et al. (2020 : 785) indiquent toutefois qu'« il faut prendre ces chiffres comme de simples indicateurs de tendance, dans la mesure où la situation peut parfois varier pour un même auteur, d'une édition à l'autre ».

-‹ier› (nous en avons expliqué l'origine dans la note de bas de p.40). Marchello-Nizia et al. (2020 : 751), citant en partie Andrieux-Reix et Baumgartner (1983), indiquent que « [l]e morphème *-z* apparaît à la P2 dans les verbes autres qu'en *-er / -ier* chaque fois que la base se termine par une consonne dentale [...]. Il sert à graphier [l'affriquée] /ts/, combinaison d'une dentale et du morphème /s/ », comme dans 2.SG *entenz*. Le -‹z› se généralise par la suite après des bases verbales à finale vocalique (autres que [ə]) comme dans 2.SG *croiz*, et on observe une alternance graphique entre -‹s› et -‹z›, notamment après la réduction des affriquées au XIIIème s (ex : 2.SG *crois* vs *croiz*). Marchello-Nizia et al. (2020 : 780) affirment que le -‹s› et le -‹t› flexionnels sont maintenus uniquement en liaison après le XIIIème s. et jusqu'en français moderne (il est question ici de l'imparfait), soit jusqu'au début du XXème s. On peut douter de cette affirmation, car les corpus enregistrés au début du siècle ne présentent pas de liaisons systématiques dans ce contexte, ce qui nous fait nous demander à partir de quand les consonnes finales de la 2.SG, mais aussi des 3.SG, 1.PL, 2.PL et 3.PL, étymologiques, sont devenues variables en liaison.

Enfin, en ce qui concerne les impératifs de 2.SG, ceux-ci proviennent directement de formes latines, qui ne présentaient pas de consonne finale (ex : lat. 2.SG AMA > fr. 2.SG *aime* ; Marchello-Nizia et al., 2020 : 775). On observe une extension analogique de -‹s› à ce temps, ce mode et cette personne pour les verbes autres que ceux en -‹er› et -‹ier›, attestée dès le XIIIème s., qui pourrait être liée selon Marchello-Nizia et al. (2020 : 775) à l'extension de cette même finale graphique à la 1.SG du présent de l'indicatif, et à nouveau, par analogie avec les 2.SG de verbes inchoatifs comme fr. 1.SG *finis*. Cette extension graphique a même parfois touché des verbes en -‹er› et -‹ier› (ex : 2.SG *parles*), mais ce n'est plus normatif en français contemporain, sauf devant les pronoms enclitiques *en* ou *y*, où une marque graphique -‹s› est apposée et où une liaison [z] est produite à l'oral (cf exemple 31 ; Marchello-Nizia et al., 2020 : 775).

Venons-en à la 3.SG. Les formes latines présentaient déjà un -[t] final à cette personne. Par la suite, pour les verbes latins en -‹are›, lorsqu'il est précédé d'un [ə] issu de la voyelle désinentielle [a] qui s'était affaiblie comme mentionné dans la note de bas de p.40, ce /t/ chute entre le IXème et le XIème s. (comme dans lat. 3.SG AMAT > fr. 3.SG *aime*) et se maintient derrière consonne pour les autres verbes, contexte découlant de la disparition des voyelles autres que [a] (comme dans lat. 3.SG VENIT > fr. 3.SG *vient* ; Marchello-Nizia et al., 2020 : 751-752). En français contemporain, on trouve tout de même pour le premier type de verbes un ‹t› euphonique lors d'une inversion sujet/verbe (cf exemple 32), alors qu'il avait préalablement chuté dans ce contexte comme je viens de le mentionner. Marchello-Nizia et al. (2020 : 763) rappellent qu'en ancien et en

moyen français, un verbe dans cette position pouvait être élidé ou non : *aim(e) il ?* ou *aime il ?*, avant que l'élision ne se raréfie à partir du XVI^{ème} s. Selon ces auteurs, la forme avec ‹t› euphonique qui se généralisera serait d'origine populaire et analogique (avec le deuxième type de verbes ; Marchello-Nizia et al., 2020 : 763).
À l'imparfait, Marchello-Nizia et al. (2020 : 787) recensent quelques attestations de chutes de -‹t› graphiques à cette personne, parfois au profit d'un -‹e› (ex : 3.SG *avoie* 'avait'), qui ne sont plus normatives aujourd'hui.
Terminons avec le passé simple, où -‹t› avait disparu pour les verbes en -‹ir› dès le XI^{ème} s., pour finir par se réimposer au XV^{ème} s. uniquement pour les verbes en -‹i› et -‹u› (Marchello-Nizia et al., 2020 : 822-823, 827).

En latin, le -[s] se retrouve aussi en position finale pour la 1.PL. En français, -‹s› n'a pas toujours été noté à l'écrit, notamment par des copistes à l'Ouest et en anglo-normand (ex : *parlon* ; Marchello-Nizia et al., 2020 : 752), mais ceci n'est plus normatif aujourd'hui.

Pour ce qui est de la 2.PL, on observe également un -[s] final sur les formes latines. Dans les terminaisons -‹atis› ou -‹etis›, le [i] a chuté, formant l'affriquée /t͡s/, notée -‹z› comme pour certaines 2.SG mentionnées plus haut. Celle-ci se réduira à /s/ dès le XII^{ème} s. au Nord (Marchello-Nizia et al., 2020 : 753), et plus généralement au XIII^{ème} s.

Enfin, à la 3.PL, le -[t] final est lui aussi présent dès le latin. Dans de rares cas, le -‹t› a pu chuter dans certains écrits (ex : *disen* ; Marchello-Nizia et al., 2020 : 763), mais encore une fois ce n'est plus normatif en français contemporain.

Nous savons donc maintenant d'où proviennent la marque de pluriel [z] et les marques de personne [z] et [t] en français. La première est issue d'une marque de cas et de nombre qui avait déjà une valeur plurielle. Pour la seconde, nous avons vu que les marques de 2.SG, 3.SG, 1.PL, 2.PL et 3.PL étaient étymologiques, et que la marque [z] avait été étendue à la 1.SG sur certains verbes au présent de l'indicatif et au passé simple, et sur tous les verbes à l'imparfait, tout comme la 2.SG a été étendue à l'impératif présent. L'analogie avec des verbes inchoatifs, avec des verbes où ‹s› appartenait à la base et y était étymologique ou l'analogie avec la 2.SG dans le premier cas sont les raisons qui sont avancées pour expliquer ces évolutions. Au contraire, la marque [t] de 3.SG a chuté sur certains verbes au présent, au passé simple et au futur de l'indicatif.

Pour retracer notre cheminement, un premier type de variation, la variation diachronique, a tout d'abord suscité mon intérêt et j'ai adopté pour commencer un angle d'approche phonétique. Après avoir rappelé à quels documents on pouvait faire appel

en phonétique diachronique pour l'étude d'états de langue n'ayant pas pu être enregistrés, nous avons passé en revue plusieurs dynamiques à l'œuvre en français, parfois contradictoires, ayant favorisé la chute ou rémergence de consonnes finales et le maintien ou la chute de consonnes en liaison, voire l'apparition de nouvelles consonnes en liaison, avant de nous pencher sur le sort de consonnes finales latines ou secondaires en français ayant pu donner lieu à des liaisons. Par la même occasion, nous avons examiné des initiales vocaliques de M2 ayant modifié leur caractère liaisonnant, pour finir par un état des lieux des connaissances sur la liaison en microdiachronie dans les corpus oraux actuels. Finalement, un détour par la morphologie diachronique aura permis de présenter certaines innovations tout en traitant de l'origine des consonnes flexionnelles de pluriel et de personne. Si cette partie était plus orientée en linguistique diachronique qu'en sociolinguistique diachronique, nous rebasculerons désormais pleinement du côté de la sociolinguistique en nous intéressant à la liaison en **diamésie**, soit la variation entre oral et écrit, avec sans plus tarder une réflexion sur le lien qui existe entre liaison et graphie.

1.2.3.2 Liaison et variation diamésique

1.2.3.2.1 Liaison et graphie

La liaison est un phénomène principalement oral, et il n'est habituellement pas indiqué à l'écrit qu'une liaison doit être ou non réalisée, même si des graphies non standard peuvent être utilisées à cet effet (dans un mail entre linguistes : exemple 1.10a, dans le nom d'un commerce : exemple 1.10b, dans le nom d'un ouvrage : exemple 1.10c, dans la correspondance de peu lettrés : exemple 1.10d (issu du corpus 14 : Praxiling - UMR 5267, 2019), dans des écrits scolaires : exemple 1.10e (issu du corpus ÉMA, écrits scolaires ; Boré et al., 2017, cf Boré & Elalouf, 2017 pour une présentation du corpus) ou encore dans un roman : exemple 26). Certaines de ces graphies sont intentionnelles (jeu de langage dans le mail entre linguistes, lettre ‹z› en filigrane du livre pour enfants, renvoyant au thème, le zéro déchet, ou encore dans le roman pour adultes, afin de mieux représenter l'oral, notamment la langue populaire), quand d'autres ne le sont pas (correspondance de Poilu, écrit scolaire).

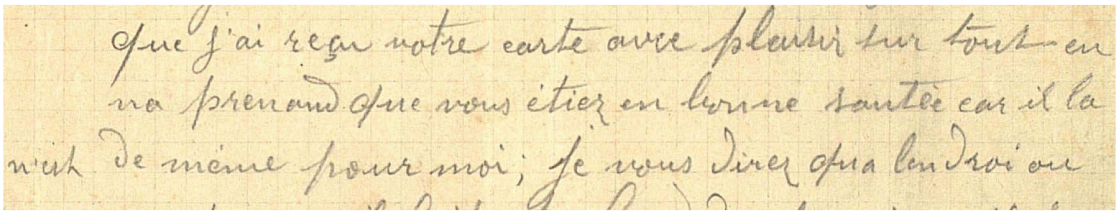
(26) – Dites voir, la Belle, chuchote le Mammouth, v'[s']avez des [z]ognons, j'suppose ?
(Frédérique Dard, San Antonio n°97 : *Si ma tante en avait*, chapitre treize, 2010)

Dans des correspondances privées de peu-lettrés du XVII[ème] et XVIII[ème] s., ayant été rédigées dans le contexte de la première vague d'expansion coloniale de la France et

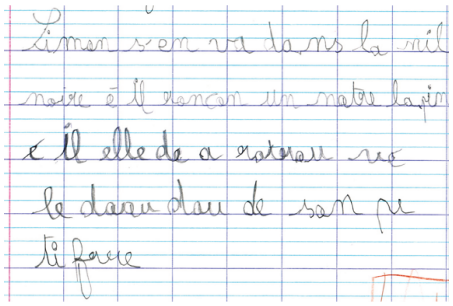
(a) *Chères zet chers*, intitulé de mail

(b) *OZOngles*, devanture d'un institut de beauté à Paris

(c) *Les Zenfants presque zéro déchet*, titre d'un livre (Pichon, J. et Moret, B. : 2016. Vergèze : Thierry Souccar Editions)



(d) « [...] *que j'ai reçu votre carte avec plaisir sur tout en na prenand que vous étiez en bonne santèe car il la n'est de méme pour moi; je vous direz qua lendroi ou* [...] », extrait d'une correspondance de Poilu (Laurent Pouchet, 01-10-1914, Corpus 14)



(e) « *Simon s'en va dans la nil noire é il rancon un natre lapin é il elle de a ratrou vé le daadau de son pe ti fare* », écrit scolaire (D2-E5, CP, 2016, Corpus ÉMA, écrits scolaires)

FIGURE 1.10 : Exemples de graphies non standard indiquant la réalisation de liaisons

faisant partie des « Prize papers », des documents saisis suite à la capture de navires marchands lors de guerres de course, Bergeron-Maguire (2021) recense trois manières de noter la réalisation d'une liaison : à la fois en finale de M1 et en initiale de M2 (exemple 27), uniquement en initiale de M2 (exemple 28) et uniquement en finale de M1, suivie d'un ‹e› graphique (exemple 29). La collection d'exemples de cette linguiste présente néanmoins un quatrième cas de figure, quand M1 et M2 sont accolés et que la consonne en question est redoublée afin d'indiquer la réalisation d'une liaison (exemple 30).

(27) *nou savon bien nan vis dan navoier des nouvell* (Louis Laniez, 1793, Pointe-à-Pitre, à sa femme au Havre)

(28) *a quelque zamis qui ont esté plus sansible* (Carré Dupaty, 1747, La Rochelle, à son frère à Saint-Domingue)

(29) *Ses Line vous dirom monne ariuest en se lieux* (P. Chevallie, 1671, La Martinique, à sa femme à La Tremblade)

(30) *vous prie et mois aussie dannavoir soin* (Elisabeth Jacob, 1670, La Tremblade, à son mari à Saint-Christophe (auj. Saint-Christophe-et-Niévès))

L'ensemble de ces exemples a été présenté dans Coutanson & Badin, sous presse.

Les liaisons peuvent apparaître après les consonnes graphiques finales suivantes (muettes lorsque les mots sont prononcés en isolation), qui peuvent donner une indication quant à la liaison possible (voir par exemple Léon, 1992 : 239) :

- /z/ : après ‹s›, ‹x› et ‹z›.
- /n/ : après ‹n›.
- /t/ : après ‹t› et ‹d›.
- //ʁ// : après ‹r›.
- /p/ : après ‹p›.

Pourtant, leur présence ne veut pas dire que le lecteur sait quand produire des liaisons, car sans connaissance des habitudes de prononciation de celles-ci, la tâche reste ardue. Par exemple, les consonnes de liaison qui se prononcent conformément à l'orthographe (voir aussi Fouché, 1956 : 436) sont ‹z›, ‹n›, ‹t›, ‹r› et ‹p›, auxquelles on peut ajouter ‹c› dans les mots composés et les locutions adverbiales, et celles dont la prononciation diffère de la graphie sont ‹s›, ‹x› et ‹d›, auxquelles on peut ajouter ‹g› dans les toponymes. Historiquement, certaines consonnes ont pu être rétablies par analogie avec des mots dérivés ou dans une visée étymologisante, comme AF *grant* > *grand* (lat. GRANDEM ; Hornsby, 2020 : 47). Il est aussi important de rappeler qu'au départ, -‹x› (final) est une abréviation de scribes pour noter -‹us›, avec ‹u› issu de la vocalisation d'un /l/ (Laborderie, 1994 ; Perret, 1998). -‹x› a aussi été rétabli là où il était étymologique (comme AF *nois* > *noix* puisque lat. NUX) car il avait pris une fonction discriminante pour indiquer que la voyelle finale de certains mots se prononçait comme une monophtongue (*ennuyeux*, *noix*, *chevaux*… ; Perret, 1998). Enfin, -‹z› (final) était au départ lui aussi une abréviation utilisée au Moyen-Âge pour noter un /ts/ à la finale, qui s'est répandu par analogie à d'autres mots, puis qui a été conservé pour les finales en /e/ car ‹z› empêchait une lecture de ‹e› comme /ə/ (Perret, 1998). Après l'introduction de l'accent aigu, le ‹z› n'est conservé que pour les mots où il n'est pas une marque de pluriel (*chez*, *nez*, *aimez*… ; Perret, 1998). Il reste cependant nécessaire pour le lecteur de savoir quelles liaisons sont invariablement réalisées et quelles liaisons au contraire ne

sont habituellement jamais faites, tout comme il doit avoir connaissance des situations socio-stylistiques dans lesquelles émergent généralement telle ou telle liaison.

Deux autres cas particuliers de liaison sont explicitement notés en français contemporain. Il s'agit tout d'abord des formes graphiques des verbes en -‹er› et -‹ier› ou irréguliers à la 2.SG de l'impératif présent devant pronom enclitique, qui présentent une consonne graphique de liaison là où les formes verbales graphiques dans un autre contexte (devant initiale consonantique, en isolation) n'en prendraient pas (exemple 31). Il en va de même en cas d'inversion sujet/verbe à la troisième personne du singulier (3.SG) au passé simple, au présent et au futur de l'indicatif (exemple 32). Cela concerne tous les verbes ne prenant pas de -‹t› ou -‹d› finals à la 3.SG. Pour le premier cas, on parle parfois de ***‹s› euphonique*** et pour le second de ***‹t› euphonique***. Trask (1996 : 134) donne la définition suivante d'***euphonie*** : "[a]ttractiveness of sound". On voit donc que ce terme relève de la phonétique impressive et réfrène notre envie de le reprendre.

(31) *mange**s**-en !*, mange**s**-y ![43] vs *mange !*, *mange-les !*

(32) *mangea-**t**-il/elle/on* ?, *mange-**t**-il/elle/on* ?, *mangera-**t**-il/elle/on* ? vs *il/elle/on mangea*, *il/elle/on mange*, *il/elle/on mangera*.

Notons au passage qu'il ne faut pas confondre ces ‹t› euphoniques avec le pronom réfléchi élidé de 2.SG *t'*. Pour ce dernier, la ponctuation différera. On écrira *va-t'en*, car il s'agit du verbe s'*en aller*. Relevons tout de même que de telles graphies non standard dans ce contexte sont anciennes, puisqu'elles étaient déjà condamnées par Vaugelas au XVII^{ème} s. (Marchello-Nizia et al., 2020 : 763). D'autres graphies inattendues peuvent apparaître pour les cas de pataquès, mais ceci sera abordé en 1.3.6.2.

Nous avons vu que la liaison est un phénomène avant tout oral et que si certaines consonnes graphiques donnent des renseignements quant aux consonnes de liaisons potentielles (j'ai inventorié les consonnes graphiques correspondant à chaque consonne de liaison), rien n'indique, à l'exception de quelques graphies non standard, que les liaisons doivent être ou non réalisées. Il n'y a pas toujours correspondance entre graphie et phonie, car certaines consonnes ont été rétablies dans une visée étymologisante quand d'autres proviennent d'abréviations. Deux cas particuliers où une consonne graphique n'est présente qu'en contexte de liaison ont été relevés, dans deux contextes verbe + enclitique : pour les impératifs de 2.SG devant *en* et *y* et pour les inversions sujet/verbe de 3.SG devant *il*, *elle* et *on*. Il est temps désormais de se consacrer à un autre type de variation, la variation **diatopique**, ou variation géographique.

[43] régional

1.2.3.3 Liaison et variation diatopique

C'est grâce à la constitution du corpus PFC, avec des enregistrements de locuteurs issus de toute la francophonie, qu'une réelle étude diatopique de la liaison a pu voir le jour. Selon Côté (2017 : 15,17), les différentes variétés du français auraient un fonctionnement similaire au niveau des liaisons invariables mais différeraient au niveau des liaisons variables. Ce que je nomme *pataquès* est également sujet à variation, mais ceci sera abordé dans la partie 1.3.6.3.

Côté (2017 : 16) nous présente néanmoins quelques exceptions parmi les contextes de liaisons invariables pour tous les locuteurs francophones, qui ne s'observeraient que dans la catégorie proclitique + proclitique/verbe :

- Comme mentionné plus tard en 1.3.2.3, *ils* ne fait pas liaison dans certains parlers de l'ouest et du nord de la France, dans les **variétés canadiennes** (français laurentien, français acadien), le français d'Old Mines dans le Missouri, tout comme en jersiais.
- On peut également rencontrer des cas où la liaison n'est pas faite après *on* dans les **variétés américaines** (variétés canadiennes + louisianaise). Dans d'autres variétés, cela pourrait passer pour des « erreurs de performance », comme par exemple dans l'étude d'un locuteur parisien par Anita Berit Hansen (2016 : 131) : celle-ci note deux non-réalisations de la liaison après *on*.
- Enfin, dans la variété louisianaise, la liaison après le clitique objet *les* n'est pas systématique.

En Louisiane, dans le français de la paroisse Lafourche, une alternance entre liaison en [z], consonne [h] et non-liaison a enfin été notée (*les Indiens* : [lezɛ̃djɛ̃~lehɛ̃djɛ̃~leɛ̃djɛ̃] ; Boutin & Lyche, 2014 : 304).

Les liaisons variables seraient quant à elles plus ou moins réalisées en diatopie (Côté, 2017 : 18) : Louisiane (0,5% à Golden Meadow) < Afrique subsaharienne (15,2%, pour les locuteurs lecteurs seulement)[44] < Canada (19,6% pour le français laurentien), Europe (≈ 20%).

En s'intéressant à des contextes lexicaux spécifiques (Côté, 2017 : 19-20), d'autres particularités se laisseraient voir, comme après les formes du verbe *être* conjuguées. Par exemple, dans les variétés canadiennes, les formes à l'imparfait ne seraient pas liaisonnantes : *est* : 75,5% de réalisations (247/327), *sont* : 66,7% (26/39) et *suis* : 28,1% (18/64) vs *étais* (0/22) et *était* (0/72) : 0%. Pour ce qui est des liaisons après des monosyllabes (cf Côté, 2017 : 19), la conjonction *quand* serait liaisonnée de façon moins

[44] La variation serait importante entre lecteurs et non-lecteurs.

systématique en Afrique : 48,2% (79/164) vs en Île-de-France : 81,7% (58/71), en Suisse 97,1% (67/69) et au Canada : 83,3% (65/78).

Les données en diatopie remettent également en cause la part de réalisation de chaque consonne de liaison dans l'ensemble des liaisons réalisées qui avaient été observées pour PFC. Par exemple (voir figure 1.4), en français centrafricain (données de Bangui), l'ordonnancement est le suivant : /z/ > /t/ > /n/ (Bordal, 2012 : 37) et en français suisse (données de Neuchâtel) : /z/ = /n/ > /t/ (Racine & Andreassen, 2012 : 199) vs /z/ > /n/ > /t/ pour l'ensemble de PFC (Durand et al., 2011 : 124). Par ailleurs, toutes les variétés ne présentent pas l'intégralité des consonnes de liaison observées, soit /p, t, n, z, ʁ/. Certaines variétés sont notamment dépourvues de liaisons en //ʁ// et /p/, comme en français albertain (Walker, 2012 : 358).

Enfin, travailler sur des données diatopiques en matière de liaison requiert une grande prudence car selon les variétés, une consonne peut être fixe ou de liaison (revoir la définition de la liaison de Durand & Lyche donnée en 1.2.1). Par exemple, Côté (2017 : 265) rapporte le cas de certains *tout* ou *fait* qu'il ne faudrait pas considérer comme des mots liaisonnants en français laurentien.

Ainsi en règle générale, les liaisons invariables sont produites sans fluctuer dans toute la francophonie, malgré quelques exceptions (comme avec *ils/on/les* + proclitique/verbe dans certaines variétés). Les liaisons variables par contre ont des taux de réalisation très inégaux en diatopie, certaines constructions (après les différentes formes du verbe *être* conjugué ou après la conjonction *quand*...) sont moins liaisonnantes dans certaines zones géographiques, et les consonnes de liaison potentielles comme leur ordonnancement peuvent différer dans les diverses zones. Les mots liaisonnants ne sont pour finir pas les mêmes dans toutes les variétés, une consonne pouvant être fixe dans l'une et de liaison dans une autre. Après avoir étudié la liaison en diachronie, en diamésie et en diatopie, laissons place à un autre type de variation, la **variation diaphasique**, qui concerne la variation situationnelle ou stylistique. Nous nous demanderons d'abord si la lecture favorise la réalisation de liaisons.

1.2.3.4 Liaison et variation diaphasique

1.2.3.4.1 Liaison et lecture

Avant les études sur corpus et expérimentales de la liaison, ce phénomène semblait plus réalisé en lecture qu'en conversation (cf par exemple Martinon, 1913 : 356), ce que les trois études présentées ci-après viendront confirmer. Ici, la variation diamésique entre

aussi en jeu, puisque la lecture à voix haute implique une oralisation de l'écrit.

Le protocole de recueil du corpus PFC prévoit quatre tâches : deux entretiens (une conversation libre tout d'abord suivie d'une conversation guidée) et deux tâches de lecture (d'une liste de mots puis d'un même texte ; Durand et al., 2002). En comparant les productions en entretien à celles qu'on obtient lors de la lecture du texte[45], Durand et al. (2011 : 113) ont pu confirmer que les liaisons étaient davantage faites en lecture (59,4%) qu'en conversation (43,4%).

Dans une autre étude portant sur des lycéens présentée en 1.2.4.1, Hornsby (2019 : 582) a noté que les liaisons variables[46] cette fois-ci étaient davantage réalisées en lecture (*Reading Style* ; 32,6% (1694/5191)) qu'en situation d'entretien[47] (*Interview Style* ; 19,9% (671/3366)). La différence s'observe au niveau de constructions spécifiques, comme dans *pas* + M2 (3/248 en entretien (1%) vs 74/188 en lecture (39%) ; Hornsby, 2019 : 584).

Enfin, dans un corpus de 32 livres audio pour enfants d'une durée totale de 17h 35min (122 363 mots et 7 368 contextes de liaison), Pustka (2017) observe davantage de réalisations de liaisons variables dans certaines constructions spécifiques que dans les entretiens du corpus PFC : par exemple, après les formes conjuguées du verbe *être* (comme après *c'est* : 87% (n = 143) dans les livres audio pour enfants vs 28% (n = 1470) dans PFC), après des adverbes monosyllabiques (comme après *bien* : 90% (n = 79) dans les livres audio vs 43% (n = 185) dans PFC) ou après des conjonctions monosyllabiques (comme après *mais* : 34% (n = 256) dans les livres audio vs 0,5% (n = 791) dans PFC).

Ainsi, le corpus PFC ayant été conçu pour comparer des conversations à des tâches de lecture, cela a permis de constater que les locuteurs liaisonnaient davantage en lecture à voix haute, ce que l'étude de lycéens de Lille, Strasbourg, Perpignan et Mons a confirmé, tout comme le corpus de livres audio pour enfants. Cette différence est visible notamment dans certains contextes variables spécifiques. Malgré les indices graphiques à disposition lors de l'oralisation d'un texte, toutes les liaisons ne sont pourtant pas réalisées. Que savons-nous concernant la liaison dans des textes versifiés ?

[45] Précisons que les mots de la liste commune à toutes les enquêtes, principalement en isolation, sont à exclure de toute étude concernant la liaison, ne comprenant pas de contextes de liaison.

[46] Dans le cadre de cette étude, Hornsby (2019 : 584) a considéré uniquement les contextes déterminant + nom et clitique + verbe comme des contextes invariables.

[47] Hornsby précise que « [t]he ninety-six informants were interviewed individually on general topics relating to their studies, interests and opinions of life in their city and region, for around twenty minutes » (Hornsby, 2020 : 148).

1.2.3.4.2 Liaison et versification

Que nous disent les règles de versification classiques à propos de la liaison ? Il faut tout d'abord noter que ce point n'est pas traité par tous les ouvrages, contrairement à la question du schwa, qui l'est systématiquement. Deux règles générales feraient globalement consensus : les liaisons ne se feraient pas entre les vers, mais seraient toutes réalisées à l'intérieur de ceux-ci (Milner & Regnault, 1987 : 52 ; Aquien, 1993 : 170).

Pourtant, tout n'est pas si simple. Milner et Regnault (1987 : 53) remarquent par exemple que la première règle est problématique en cas d'enjambement, ce qu'ils illustrent à l'aide des exemples 33 et 34 :

(33) Rome, aussi bien que moi, vous donne son suffrage,
Répudie Octavie, et me fait dénouer
Un hymen que le ciel ne veut point avouer (Racine, *Britannicus*, II, 3)

(34) … Hé quoi ? N'avez-vous pas
Ordonné dès tantôt qu'on observe ses pas ? (Racine, *Bérénice*, IV, 6)

Aquien propose l'exemple suivant pour un autre cas problématique relatif cette fois-ci à la seconde règle, à savoir la présence d'un signe de ponctuation en vers et en contexte de liaison :

(35) Tu disais ; et nos cœurs unissaient leurs soupirs (Lamartine, *Méditations poétiques*, L'immortalité)[48].

Enfin, un dernier cas problématique relatif à la seconde règle est la réalisation de la liaison à la césure. Milner et Regnault (1987 : 54-55) présentent des exemples tels que les suivants :

(36) Voici tous mes forfaits. En voici le salaire (Racine, *Britannicus*, IV, 2)

(37) Qu'ils semblent tous venger un père comme vous (Corneille, *Cinna*, I, 3)

Pour ces trois cas, difficile de savoir si la liaison doit être ou non réalisée. Dans leur ouvrage *Dire le vers*, Milner et Regnault (1987 : 55-56) proposent de réserver la diction directe pour les cas de liaison problématiques, c'est-à-dire de produire des liaisons non enchaînées. De cette façon, la réalisation de liaisons non enchaînées a pu être conseillée pour des cas de liaisons problématiques en récitation de vers. Mazaleyrat (1974 : 148) au contraire propose d'enchaîner les liaisons à la césure. Selon lui, « […] la nature de la césure, ce n'est pas d'être un arrêt. Elle coïncide avec un arrêt si la grammaire et

[48] Notons que selon les éditions, le signe de ponctuation de ce vers peut différer ou être absent.

le sens le demandent. Mais ce n'est pas cet arrêt qui la constitue » (Mazaleyrat, 1974 : 148-149). On voit donc que diverses propositions sont apportées pour traiter ces cas problématiques.
D'autres indications mineures peuvent être apportées : par exemple, Aquien (1993 : 170) indique qu'« il est d'usage de ne faire sentir aucune liaison en cas d'interruption, de changement d'interlocuteur, et devant les interjections ».

Une règle supplémentaire revient régulièrement à propos de la liaison dans les traités de versification : la **règle de la liaison supposée**, qui interdit d'opter pour deux mots à la rime s'ils ne liaisonnent pas de la même façon (Aquien, 1993 : 241). Par exemple, *flanc* pourrait rimer avec *sang* (on parle de ***consonnes équivalentes*** pour ‹c› et ‹g› ici ; Mazaleyrat, 1974 : 200) mais pas avec *ans*. Mazaleyrat (1974 : 200-201) apporte l'explication suivante :

> [...] elle est un vestige de l'histoire. Les habitudes en matière de rime se sont établies en un temps où beaucoup de consonnes finales étaient perceptibles à la pause, donc aussi en fin de vers quand chacun formait une unité distincte d'élocution ; où les sonores s'y assourdissaient [...]. C'est sous ce point de vue que la rime pour l'œil sur consonnes finales représente la conversion d'un état phonétique ancien en règle de l'art.

Abordons un dernier point, très important en ce qui nous concerne. Delattre (1955 : 45) indique pour la phrase : *Des hommes illustres ont attendus* que « [d]ans la récitation de vers, il serait de rigueur de lier l'*s* d'*illustres*, ce qui ajouterait une syllabe phonétique à la phrase ». Pourtant, dans ce cas ce n'est pas vraiment la liaison qui a une influence sur le nombre de syllabes mais plutôt le sort du schwa qui importera. Il pourrait sembler étonnant que ce dernier soit réalisé sans liaison, mais il est impossible d'affirmer pour autant que l'ajout d'un schwa dans ce contexte entraîne automatiquement l'ajout d'une liaison. L'exemple 38 issu du FJD en est un exemple flagrant : le schwa à la fin du mot *jeunes* est réalisé, sans que la liaison ne le soit. Ainsi, selon moi, il ne faut pas faire appel à l'argument de la métrique pour expliquer le phénomène de pataquès.

(38) *SGR : toi les deu(x) jeunes et moi les deu(x) grands.
(JD_a287_06, 63550, Marie Fontenille)

En conclusion, les liaisons ne seraient pas produites entre les vers mais elles seraient toutes faites à l'intérieur d'un même vers. Il est pourtant difficile de savoir comment traiter en contexte de liaison les cas d'enjambement, la ponctuation interne au vers ou

encore les césures : faut-il dans ces contextes ne pas faire les liaisons, les faire sans enchaînement ou avec enchaînement ? Dans ce point, il a également été question de la règle de la liaison supposée qui impose de choisir à la rime des mots liaisonnant semblablement. Enfin, contrairement à une idée assez répandue, la (non-)réalisation d'une liaison ou d'un pataquès ne bouleverse pas, selon moi, la métrique d'un poème ou d'une chanson : ce serait plutôt le comportement de schwas qui en serait la cause. Puisque qu'elles sont mentionnées, que peut-on dire de la liaison dans les chansons ?

1.2.3.4.3 Liaison et chanson

Il n'existe jusqu'à présent, à ma connaissance, qu'une étude portant sur la liaison dans les chansons. Il s'agit de celle de Nardy et al. (2014), qui s'intéressent aux enfantines, ou « genres oraux utilisés par les enfants au cours de leurs jeux chantés : chansons, formulettes, comptines, devinettes, charades, etc. » (Chauvin & Colletta, 2002 : 40). Dans le cadre de sa thèse de doctorat, Chauvin (1999) a effectué une enquête de terrain de type ethnographique d'une durée de quatre ans dans les cours de récréation d'écoles primaires de l'ancienne région Rhône-Alpes, prenant la forme d'une observation directe, où elle a enregistré et filmé des comptines relevant de la tradition orale enfantine, émises spontanément par des écoliers âgés de 6 à 12 ans, individuellement ou collectivement. Nardy et al. (2014) ont repris ces données et ont conservé 36 enfantines sur les 92 recueillies, dont le texte présentait au moins un contexte de liaison. Dans celles-ci, le taux global de réalisation des liaisons atteint 74% (348/473) et celui des liaisons variables 59% (177/299) ; Nardy et al., 2014 : 251-252). Les auteurs n'observent presque pas de variation entre les différentes versions d'une même enfantine, à l'exception du couple M1-M2 *vais à* : seize autres contextes sont toujours liaisonnants, et six autres ne le sont jamais (Nardy et al., 2014 : 252-253). Les liaisons dans les enfantines semblent donc apprises en même temps que le texte, la mélodie, la diction voire une gestuelle (Nardy et al., 2014 : 256). Ces linguistes finissent par se demander dans quelle mesure les liaisons réalisées en parole chantée peuvent influencer la parole ordinaire : « [n]ous pensons qu'une séquence des enfantines incluant un contexte de liaison peut fournir le matériau constitutif d'un schéma à la condition que cette séquence soit aussi suffisamment fréquente dans les échanges quotidiens » (Nardy et al., 2014 : 258).

Ainsi, un seul corpus d'étude de la liaison en chanson existait jusqu'alors : un corpus d'enfantines. Les liaisons s'y révèlent fortement réalisées, avec une variation quasi-inexistante entre les différentes versions d'une même enfantine. On peut se demander si la façon de liaisonner des chanteurs peut affecter leurs productions quotidiennes en

tant que locuteurs.

Nous avons abordé la liaison sous l'angle de la variation diaphasique en lecture, puis dans les textes versifiés, et enfin dans les chansons. Pour l'heure, nous porterons notre attention sur la **variation diastratique**, c'est-à-dire la variation qui existe entre les différents groupes sociaux.

1.2.4 Liaison et variation diastratique

1.2.4.1 Liaison et différences sociolectales

Dans une étude variationniste de Hornsby parue en 2012 (pour un travail de terrain réalisé entre 1998 et 2003), un échantillon de 48 lycéens de Lille et Strasbourg en classe de première ou terminale (16-19 ans), équilibré filles/garçons, s'est vu proposer une tâche de lecture d'un texte spécifiquement conçu pour présenter un grand nombre de contextes de liaisons (Hornsby, 2012 : 71-72, 2020 : 148)[49]. Les lycéens étudiaient soit en lycée général (lycée proposant par ailleurs une classe préparatoire aux grandes écoles) ou en lycée professionnel (Hornsby, 2012 : 72). En moyenne, les premiers lycéens appartiendraient à des classes socio-économiques plus élevées que les seconds (Hornsby, 2012 : 72). Dans cette étude, les adolescents issus de milieux plus favorisés, étudiant en lycée général, présentent un taux de réalisation des liaisons en lecture supérieur aux adolescents issus de milieux moins favorisés, étudiant en lycée professionnel (36,4% (459/1260) vs 30,3% (385/1269)[50] ; Hornsby, 2012 : 73). Selon le linguiste, « [c]ontrol of optional liaison would seem to be a product of what Bourdieu (1982 : 36) would term *capital linguistique* [...] acquired within the home or via the education system [...] » et « [...] well before they enter the workplace, those on whom the prestige of the written word is more likely to confer power and authority are significantly more able to navigate one of its most notorious pitfalls » : la liaison variable (Hornsby, 2012 : 79). La différence observée en lecture ne semble pas se retrouver en conversation : en effet, les lycéens de milieux favorisés liaisonneraient autant dans ce contexte (20,2%) que les lycéens de milieux moins favorisés (19,6% ; Hornsby, 2020 : 163). Il y a donc ici intersection entre variation diastratique et variation diamésique, puisqu'une différence

[49] Dès 2019, Hornsby présente des résultats incluant également les villes de Perpignan et Mons, avec une fourchette d'âge pour dorénavant 96 enquêtés allant de 15 à 20 ans (Hornsby, 2019 : 582, 2020 : 148).

[50] Il m'a été impossible de retrouver les pourcentages présentés dans Hornsby, 2012 : 73. J'ai pour ma part recherché le nombre de liaisons correctes et incorrectes réalisées sur le nombre de contextes de liaison (c'est-à-dire en excluant les données manquantes).

ne s'observe qu'en lecture.

Pour résumer, Hornsby (2012) a montré que les lycéens issus de milieux favorisés réalisaient plus de liaisons en lecture que les lycéens issus de milieux moins favorisés, bien que dans leurs conversations quotidiennes, ils convergeraient vers des taux similaires. La scolarité de ces lycéens diffère déjà. Certains poursuivront encore leurs études dans l'enseignement supérieur. Dans la foulée, nous aborderons maintenant la question de la réalisation des liaisons selon le niveau d'étude.

1.2.4.2 Liaison et différences de niveau d'étude

Encrevé (1983 : 47) parlait pour la liaison de « phénomène sociolinguistique inversé », car : « [t]outes les données connues et toute observation directe indiquent que ce sont les locuteurs du français les plus scolarisés qui présentent le plus large système de variation sur la liaison ». Pourtant, Durand, Laks et al. (2011 : 128) ne notent pas de réelle différence entre les locuteurs en fonction de leur nombre d'années d'étude, dans la conversation ordinaire. Ces auteurs émettent cependant l'hypothèse que les locuteurs, dans une situation de forte tension linguistique telle une prise de parole publique, auraient pu présenter des comportements différents (Durand et al., 2011 : 128). Pour la moitié des locuteurs environ, l'information n'avait en outre pas été renseignée, excluant les locuteurs concernés de cette analyse, et les linguistes admettent enfin qu'un focus sur les liaisons variables pourrait amener à nuancer ces observations (Durand et al., 2011 : 128). Une dernière critique pourrait être faite concernant cette étude, à propos de la subdivision des locuteurs en trois groupes : ceux avec moins de quatorze années d'études (ce qui correspond à des locuteurs qui n'auraient pas terminé leurs études secondaires), ceux ayant été scolarisés entre quatorze et vingt ans au total (ce qui englobe les locuteurs des titulaires du baccalauréat aux titulaires d'un diplôme de master) et ceux ayant étudié plus de vingt ans (ce qui correspond à des locuteurs ayant un diplôme supérieur à bac + 5). Cette subdivision pose problème. D'abord bien que le critère de définition du premier groupe semble justifié, ce dernier ne comporte qu'un locuteur, ce qui est insuffisant pour une comparaison inter-groupe. Quant aux deux autres groupes, il aurait été plus judicieux de subdiviser le premier, trop large, avec par exemple d'un côté les locuteurs ayant un niveau d'étude compris entre bac et bac + 2 et de l'autre ceux ayant un niveau d'étude compris entre bac + 3 et bac + 4, en faisant basculer ceux ayant atteint le niveau bac + 5 au moins au dernier groupe. Si on reprend l'affirmation d'Encrevé située en début de paragraphe, elle semble à nuancer : il y a

tout lieu de penser, en fonction des données de PFC, qu'entre la conversation ordinaire des locuteurs les plus éduqués et celle de locuteurs moins éduqués, on n'observe pas de différence significative, même si la situation pourrait avoir évolué depuis lors. Pour autant, dans le cadre de leur travail, par exemple dans la parole politique étudiée chez Encrevé (1988), ou dans d'autres situations à enjeu social élevé, le comportement des locuteurs pourrait être tout autre. Ici, il y a donc une intersection de facteurs de variation (variation diastratique et variation diaphasique).

Pour conclure brièvement, dans PFC, de premiers résultats n'ont donné à voir aucune différence de réalisation au niveau des liaisons, invariables et variables confondues, en fonction du niveau d'étude des locuteurs. Ceci est à prendre avec précaution, de nombreuses données faisant défaut et les délimitations des groupes étant peut-être à rediscuter. De la variation pourrait être révélée en favorisant l'étude des liaisons variables ou si les conditions de recueil présentaient davantage de tension linguistique.

Nous avons donc convoqué les différentes dimensions de la variation afin de voir ce que chacune avait à nous apprendre sur le phénomène de liaison, tout en relevant souvent qu'elles se combinaient entre elles. La dimension diachronique tout d'abord nous a amenés à faire preuve de prudence, le phénomène étant bien plus complexe qu'une simple chute progressive d'anciennes consonnes finales. Par la suite, la dimension diamésique nous a montré qu'un phénomène oral comme la liaison entretenait des rapports complexes avec l'écrit. Pour continuer, le rôle de la dimension diatopique s'est révélé assez limité dans le cas de la liaison. La dimension diaphasique pour sa part nous a montré que certains styles, comme la lecture à voix haute, étaient plus liaisonnants. Quant à la dimension diastratique, si une différence s'observait en lecture seulement chez les lycéens issus de milieux favorisés, il n'y avait pas de différence en conversation selon cette fois-ci le niveau d'étude des locuteurs.

Nous avons à présent une très bonne image d'ensemble du phénomène de liaison, que nous avons décortiqué d'abord dans une perspective linguistique, puis sociolinguistique. Dès lors, il est temps de proposer un état de l'art sur un phénomène rare qui existe en marge du phénomène de liaison, le pataquès, et qui m'intéresse tout particulièrement. Je commencerai par présenter une réflexion quant à l'appellation que j'ai retenue pour ce phénomène.

1.3 Le pataquès : un cas particulier de liaison

1.3.1 Le pataquès : à la recherche d'une appellation

On trouve dans la littérature spécialisée une majorité de termes dépréciatifs pour se référer à notre objet d'étude[51] : *liaisons abusives* (Blanche-Benveniste, 1997 : 152), *erreurs de liaison* (Desrochers, 1994 : 243), *liaisons erronées* (Desrochers, 1994 : 249), *incorrect liaisons* (Armstrong & Pooley, 2010 : 115), *liaisons fausses/fausses liaisons/false liaisons* (Bauche, 1920 : 58 ; Frei, 1929 : 188 ; Gaatone, 1979 : 328 ; Tranel, 1987 : 169 ; Gadet, 1992 : 43, 1989 : 56 ; Desrochers, 1994 : 243 ; Hornsby, 2012 : 74 ; Marchello-Nizia et al., 2020 : 486), *fautes de liaison* (Gaatone, 1979 : 328 ; Gadet, 1989 : 55), *liaisons fautives* (Gaatone, 1978 : 51 ; Léon, 1984 : 1 ; Léon, 1992 : 247 ; Gadet, 1989 : 56 ; Nardy et al., 2014 : 246 ; Hutin et al., 2022), *(liaisons) mal t'à propos* (Encrevé, 1983 : 48 ; Léon, 1984 : 3), terme qui a néanmoins le mérite d'illustrer immédiatement le phénomène en question avec un premier exemple, *liaisons malencontreuses* (Léon, 1984 : 5) ou encore *liaisons vicieuses* (Gaatone, 1979 : 330 ; Béchade, 1992 : 53). Les consonnes impliquées peuvent aussi être qualifiées de *parasites* ou *parasitiques* (Müller, 1985 : 174 ; Léon, 1992 : 245 ; Desrochers, 1994 : 248). Ces termes sont très problématiques car ils relèvent de la prescription, position que le linguiste évitera au profit d'une approche descriptive. D'autres termes n'ont pas forcément de connotation dépréciative, comme celui de *liaisons dangereuses* utilisé par Desrochers (1994 : 243)[52].

Les termes *cuirs*, *velours* et *pataquès*, qui sont des hyponymes, ne présentent pas pour leur part un caractère dépréciatif marqué au niveau lexical. Malgré des divergences dans les définitions traditionnelles proposées, on peut tout de même retenir que le terme ***velours*** est souvent associé à l'insertion d'un [z], ***cuir*** à l'insertion d'un [t] et *pataquès* à la permutation d'une consonne de liaison avec une autre (Martinon, 1913 : 359). Sans être précisé, il va de soi que ces insertions ou permutations se produisent en contexte de liaison. Pour ma part, j'ai opté pour le terme *pataquès* pour nommer le phénomène que j'étudie. Ce que les non-spécialistes nomment *pataquès* (comme nous venons juste de le voir) a une acception généralement moins large que celle qui sera utilisée au sein de ce travail de recherche.

[51] Éloignons ici les termes entre guillemets, qui semblent pouvoir indiquer une reprise du terme à un autre auteur (avec acceptation ou réfutation du terme) ou une volonté de minimiser leur connotation dépréciative et prescriptive.

[52] Il semble surtout y avoir un aspect humoristique ici avec un clin d'oeil à l'œuvre de Pierre Choderlos de Laclos. Ce terme est problématique car il est souvent utilisé dans cette optique dans les titres d'articles de domaines scientifiques très variés. Ainsi, il n'est suffisamment précis pour être adopté.

Nous avons vu que de nombreux termes dépréciatifs ont jusque là été employés pour se référer à notre objet d'étude. Si j'ai adopté le terme *pataquès*, c'est en y faisant appel en tant qu'hyperonyme. À présent, il est nécessaire de préciser en quoi consiste le phénomène sur lequel mon regard s'est penché.

1.3.2 Le pataquès : à la recherche d'une définition

Quelle définition adopter pour la notion de *pataquès* ? Plutôt que de considérer ce phénomène comme extérieur au phénomène de liaison, il semblerait plus judicieux de le considérer comme en faisant partie (contrairement à des termes comme *fausse liaison* qui pourraient laisser entendre que ce ne sont pas des liaisons). En effet, les pataquès semblent provenir de cas de liaisons (nous nous pencherons plus amplement sur cette question par la suite, en 5.5). Voilà pourquoi notre définition doit inclure le terme de *liaison*.

1.3.2.1 Le pataquès : une liaison non normative ?

Tout d'abord, le pataquès semblait pouvoir être présenté comme une liaison non normative, appellation qu'on retrouve par exemple dans Blanche-Benveniste, 2004 : 140. Cela est lié à la notion de norme, pour laquelle Baylon (1996 : 162) propose deux définitions différentes : tout d'abord comme une « moyenne […], la langue sans les écarts [(les registres populaires et familiers comme les usages trop distingués)] » qui constitue une « norme d'usage, statistique, qui représente le registre des habitudes linguistiques sociales qui se régularisent par la vie en société » ou encore comme un « modèle à imiter […] : ensemble des formes habituellement considérées comme correctes ». Avec la première définition, on verrait le pataquès comme un phénomène linguistique non normatif, c'est-à-dire habituellement non réalisé dans la communauté linguistique francophone. Avec la seconde, le pataquès serait vu comme un élément linguistique non normatif, c'est-à-dire prohibé par certains membres de la communauté linguistique francophone, puristes, qui imposeraient aux autres membres une certaine façon de parler (ici sans pataquès). Lodge (2011 : 77) parlerait pour la première de ***norme descriptive*** (sans jugement de valeur) et pour la seconde de ***norme prescriptive*** (fondée sur des jugements de valeur).

Dans un premier temps, cette proposition pose des problèmes quant à la manière de vérifier quelles occurrences sont des liaisons non normatives et quelles occurrences n'en sont pas. Parler de *liaison non normative* en adoptant la première définition imposerait

une vérification statistique sur corpus. Plusieurs problèmes émergent : pour commencer, il est tout à fait probable que les pataquès soient réalisés de façon au moins aussi importante que des liaisons erratiques qui elles auraient potentiellement plus de chance d'être tolérées car conformes à la graphie. En se fondant sur la fréquence de ces deux types de liaisons, semblables, l'étiquette *liaisons non normatives* ne permettrait pas de les distinguer et on aurait ainsi une méta-catégorie comprenant des éléments non désirés dans le cadre de cette étude. Autre argument important, tous les pataquès ne sont pas réalisés avec la même fréquence. Par exemple, il existe des constructions où des pataquès sont fréquemment produits alors que certains pataquès constituent plutôt des cas isolés (voir 1.3.5.3). De cette façon, certains pataquès apparaîtront plus habituels, probables, que d'autres. Seront-ils alors à égale mesure considérés comme des liaisons non normatives ? Parler de *liaison non normative* en adoptant la seconde définition imposerait cette fois-ci une vérification notamment dans des grammaires, mais comme tous les cas ne peuvent pas y être présentés, il y a de fortes chances qu'une classification d'une occurrence comme pataquès résulte alors d'une considération introspective, ce qui est un critère tout à fait relatif. On pourrait néanmoins recourir à des jugements de grammaticalité proposés à des locuteurs francophones variés, toutefois tous les pataquès possibles devraient pouvoir être listés au préalable afin d'être testés, ce qui est impensable.

Dans un second temps, le plan diachronique creuse un peu plus le fossé entre les deux définitions possibles de cette proposition. Avec la première définition, l'évolution de la langue pourrait faire fluctuer le caractère normatif ou non des pataquès au gré des usages, étant donné qu'il s'agit d'une norme d'usage, tandis qu'avec la seconde, les jugements portés sur la langue auraient davantage tendance à ne pas évoluer aussi vite que les usages, étant donné qu'il s'agit d'une norme prescriptive. Selon la définition choisie, les occurrences seraient de la sorte plus ou moins normatives diachroniquement.

De toute façon, un ultime argument sonne le glas d'une possible utilisation de la notion de norme pour définir le pataquès, puisqu'il existe plusieurs normes, une pour chaque groupe socioculturel. On parle de norme groupale. Si le pataquès peut être vu comme non normatif dans la variété standard, c'est-à-dire selon la norme collective, est-ce le cas pour toutes les autres variétés de français ?

Si nous n'avons opposé que norme descriptive à norme prescriptive, et norme groupale à norme collective, nombre d'autres normes sont mentionnées dans la littérature sociolinguistique : norme objective vs norme subjective, norme fantasmée, norme endo-

gène vs norme exogène… En outre, selon la définition adoptée, l'analyse des occurrences comme normatives ou non présenterait des biais méthodologiques : rareté équivalente des pataquès et des liaisons erratiques qui empêcherait de catégoriser les occurrences selon leur fréquence, constructions mettant en jeu plus souvent certains pataquès et les rendant ainsi plus habituels, catégorisation par introspection, flou pour répartir les données selon la prise en compte ou non de la diachronie… Comme nous l'avons vu, en raison de ces multiples façons d'aborder la norme et de trouver des critères objectifs (critères statistiques ou utilisation de grammaires, de jugements de grammaticalité…) pour classifier chaque occurrence, la notion de non normatif sans précision complémentaire ne se révèle pas satisfaisante pour définir ce que sont les pataquès.

1.3.2.2 Le pataquès : une liaison non standard ?

La notion de standard se révèle-t-elle plus satisfaisante ? Marie-Hélène Côté (2017 : 20) fait notamment mention de liaisons non standard pour se référer à notre objet d'étude. Notons que le terme « standard » n'est au départ qu'un équivalent du mot « norme », issu de l'anglais et fréquent dans le domaine technique (Baylon, 1996 : 163). Sachant cela, les problèmes rencontrés avec la notion de norme subsistent.

Toutefois, la notion de standard est souvent utilisée pour distinguer la **variété standard** d'une langue de ses variétés non standard. La variété standard est la plus prestigieuse ; elle trouve son origine dans les habitudes linguistiques des classes privilégiées ou dirigeantes ; elle est aussi utilisée dans les domaines où le contrôle social est le plus important ; enfin, elle est souvent fixée par l'écrit, ce qui la rend plus stable par rapport aux autres variétés non standard (Ammon et al., 1987 : 275). Milroy (2006 : 134, 136, 138) ajoute que cette variété, vue comme correcte, légitime et faisant autorité, est acquise à l'école, mais que celle-ci ne reste qu'une idéalisation. Enfin, il s'agit de la norme de référence (Bulot & Blanchet, 2013 : 8).

En ce qui nous concerne, la variété standard ne comprendrait pas de pataquès alors que les variétés non standard pourraient en présenter. Bayley (2013 : 530) rappelle qu'un locuteur d'une variété non standard s'expose au jugement d'autrui et qu'il peut être vu comme ayant notamment une intelligence déficiente. Produire des pataquès est, dans cette mesure, stigmatisé.

Si l'utilisation de la notion de standard peut sembler plus précise que celle de norme, ce terme n'est pas pour autant parfaitement convaincant. Le problème majeur du concept de non standard, tout comme celui de non normatif comme vu précédemment, est que ceux-ci peuvent être fluctuants en fonction de la frange de la population

prise en considération. De qui est en effet constituée la communauté linguistique francophone ? Par exemple, diatopiquement, les locuteurs francophones à travers le monde forment un ensemble plutôt hétérogène et une variété standard locale pourrait coexister avec la variété standard hexagonale.

En résumé, on qualifie souvent de *non standard* un phénomène linguistique absent de la variété standard, ici, le pataquès. En produire serait stigmatisant. Le risque en utilisant les notions de norme ou de standard est de s'entendre dire que tout n'est pas également non normatif ou non standard pour tous les locuteurs du français.

1.3.2.3 Le pataquès : une liaison non étymologique, non historique ?

Pour éviter de faire appel aux notions de norme et de standard, deux autres définitions restent à notre disposition. Une première proposition serait de définir les pataquès comme des liaisons non étymologiques ou non historiques. C'est par exemple ce qu'emploie Yves-Charles Morin (1982) dans son article *De quelques [l] non étymologiques dans le français du Québec : notes sur les clitiques et la liaison* pour parler d'un cas très spécifique de pataquès sur lequel nous reviendrons par la suite. Cela s'inscrit dans une tradition plus ancienne avec des termes déjà existants pour caractériser de tels sandhis externes tels qu'***épithèse*** (« [a]djonction (gr. epi-thesis) en fin de mot d'un phonème non étymologique » ; CNRTL, 2012b) ou ***phonème paragogique*** (« [a]jouté à la finale d'un mot », la dimension étymologique se retrouvant dans la définition générale, davantage centrée sur l'écrit : ***paragoge*** : « [a]ddition à la finale d'un mot, d'une lettre ou d'une syllabe non étymologique » ; CNRTL, 2012i).

Pour ne pas tomber dans le travers du lien avec l'écrit, étant donné qu'une langue est d'abord parlée avant d'être écrite (pensons particulièrement aux locuteurs analphabètes ou illettrés), il pourrait être pertinent de faire appel au concept d'étymologie, non au sens de l'influence de la connaissance de l'étymologie par les locuteurs sur leur production de pataquès mais plutôt comme critère classificatoire utilisé par le chercheur distinguant les liaisons étymologiques, d'une part, et les pataquès, liaisons non étymologiques, d'autre part. Cette notion n'est pourtant pas toujours claire : quelle étymologie de référence utiliser ? Faut-il se référer aux étymons latins des M1, sachant que tout le lexique français n'est pas d'origine latine ? À des formes plus tardives, mais toujours antérieures au français contemporain ? Opter pour les étymons latins comme point de référence semblerait, au premier abord, un choix judicieux quand on sait que nombre de graphies finales muettes sont, en français, des graphies étymologisantes. Nous pren-

drons toutefois un exemple qui montrera que ce choix n'est pas le plus approprié. Le pronom sujet de 3.PL.Masculin (M) *ils* du français provient du démonstratif latin *ĭllī* > *īllī* (Marchello-Nizia et al., 2020 : 682). À l'écrit,

> [...] la forme avec *-s* final apparaît dans quelques textes de la seconde partie du 13e s. [...]. *Ils* et *ilz* deviennent fréquents à partir de 1350 [...]. Aux 16e et 17e s. plusieurs grammairiens signalent la prononciation *iz* devant voyelle, prononciation qui perdure en [français moderne] (Marchello-Nizia et al., 2020 : 682).

Le pronom sujet de troisième personne du pluriel masculin s'est ainsi longtemps réalisé *il* sans flexion devant une initiale vocalique, ce qui se retrouve encore dans le parler de certains locuteurs de corpus actuels, comme PFC :

(39) *JB : et quelques lapins s'ils ont pas la maladie.
[ekɛlkəlapɛ̃silɔ̃palamaladi]
(PFC, Domfrontais (Normandie, France, Europe), 2003, 61ajb1)

C'est aussi le cas dans l'enquête en Vendée de PFC (Pustka, 2009 : 322-324) ou dans le corpus ESLO (Dugua & Baude, 2017 : 49 ; pour le corpus, cf Eshkol-Taravella et al., 2011), et c'est même la forme standard du français d'Old Mines dans le Missouri (Thogmartin, 1970 : 47) ou en jersiais (Jones, 2012 : 220), un dialecte du normand (une langue d'oïl). Si la forme prévocalique sans flexion est [il] dans tous ces exemples, on trouve aussi la forme [j], habituelle dans les variétés canadiennes, comme dans l'exemple suivant :

(40) J'imagine qu'ils ont dû [...].
[ʒimaʒinkjɔ̃dy]
(PFC, Trois-Rivières (Canada, Amérique du Nord), 2010, cqejb1)

La production d'une glissante est aussi mentionnée pour le Nord-Pas-de-Calais (Dawson et al., 2016 : 155) et la Normandie (Buscail & Lyche, 2016 : 166). Ainsi, une liaison telle que la suivante :

(41) Mais si ils ont fixé la date ils ont dit que [...].
[mɛsiilzɔ̃fikseladatilzɔ̃dikɛ]
(PFC, Bamako (Mali, Afrique subsaharienne), 2006, maazw1)

n'est pas une liaison étymologique en prenant comme état de langue de référence le latin ou l'ancien français mais l'est en prenant un chronolecte postérieur. Ce cas est similaire à un autre cas que nos avions abordé en 1.2.2.2.4, celui des déterminants

unipossessifs féminins qui sont passés de formes prévocaliques sans liaison (comme *m'amie*) à des formes prévocaliques avec liaison (comme *mon amie*). Définir les pataquès comme des liaisons non étymologiques est quelque peu problématique quand on sait que les deux cas que je viens de mentionner comptent aujourd'hui parmi les liaisons les plus fréquentes du français. Par ailleurs, ces deux cas présentent forcément de la variation pour arriver d'une forme A à une forme B. On pourrait considérer qu'une consonne devient étymologique quand sa réalisation tend vers 100% à un état donné de la langue (comme *ils* devant initiale vocalique en français standard qui liaisonne en [z]). Une autre option serait de faire appel au concept de liaison non historique pour éviter de renvoyer inconsciemment vers des états très anciens de la langue.

Ainsi, parler de liaisons non étymologiques ou non historiques permettrait d'adopter une autre approche qui éviterait d'avoir recours aux notions de norme et de standard. Ces liaisons relèveraient alors du phénomène d'épithèse ; elles pourraient aussi être caractérisées comme des phonèmes paragogiques. L'étymologie de chaque M1 permettrait de trier les occurrences avec d'une part les liaisons et de l'autre les pataquès. Mais nous avons vu avec les exemples des liaisons après *ils* et des déterminants unipossessifs féminins, qui ne sont pas étymologiques, que la question du chronolecte de référence subsiste si on décide d'adopter ces définitions.

1.3.2.4 Le pataquès : une liaison inattendue

Une dernière proposition, le pataquès comme liaison inattendue, amène les questions suivantes : inattendue, par qui ? Par rapport à quoi ? Selon quels critères ? Comme nous allons le voir, on s'expose à revenir aux notions de norme et de standard ou encore d'étymologie. Pour autant, cette définition sera celle que j'adopterai.

Pour commencer, le pataquès pourrait constituer une liaison inattendue soit par les auditeurs, soit par le chercheur. Ensuite, son caractère inattendu pourrait prendre comme point de référence soit les productions habituelles des locuteurs (retour à la notion de norme), soit ce qui est standard dans la langue (retour à la notion de standard), soit une forme diachroniquement antérieure à la forme contemporaine (retour à la notion d'étymologie). Enfin, seuls des résultats expérimentaux (comme une mesure des temps de réaction) ou des données conversationnelles (demande de répétition de la part de l'interlocuteur, durée de la pause entre les deux tours de parole…) pourraient permettre de certifier ce caractère inattendu pour les auditeurs. Si les pataquès étaient pris comme inattendus par le chercheur cette fois-ci, cela pourrait se vérifier soit en

fonction d'un critère descriptif (avec vérification probabiliste sur corpus), soit selon un critère normatif (vérification à l'aide de grammaires ou de jugements de grammaticalité), encore faudrait-il comparer le caractère inattendu des pataquès à celui de liaisons erratiques, soit selon un critère étymologique.

Je prendrai les **pataquès** comme des liaisons inattendues par rapport à la consonne graphique finale du M1[53], le critère retenu, un dernier critère possible, sera donc orthographique. Ce n'est pas forcément le choix le plus pertinent, car cela nécessite par exemple d'inclure dans la catégorie pataquès les liaisons qui apparaissent dans les inversions sujet-verbe (comme en 42) et qui sont pourtant unanimement considérées comme normatives, mais c'est celui qui nous permettra de relever les occurrences les plus indiscutables, car, il faut bien se rendre à l'évidence, c'est le critère que les puristes finissent toujours par invoquer pour se justifier et la solution pour laquelle les linguistes optent (peut-être de façon plus implicite).

(42) *SGR : a [*t] ell(e) pas bien gagné l(e) congé d(e) son bien + aimé ?
(JD_a061_13, 43810, Virginie Granouillet)

Cette définition permet de passer outre les problèmes de normes ou de variétés standard multiples. Elle présente également le mérite de ne pas renvoyer à une étape imprécise du changement linguistique. D'un point de vue méthodologique, ce choix est le plus simple car il n'est alors pas nécessaire de mettre en place des études expérimentales ou sur corpus au préalable afin de sélectionner les cas concernés, avant d'être en mesure d'annoter nos propres données. Il aurait fallu dans le second cas des corpus conversationnels massifs pour le français dont nous ne disposions pas afin de mener à bien cette étude. Enfin, soulignons que le critère graphique conservé est facilement traitable de façon informatisée.

Pour conclure, j'ai décidé de définir le pataquès comme une liaison inattendue par rapport à la consonne graphique finale du M1. Avec l'adoption de cette définition, la liaison est inattendue par le chercheur et nous n'avons besoin que de dictionnaires pour classifier les occurrences comme attendues ou inattendues. Je suis consciente des limites de cette proposition, par exemple les ‹t› euphoniques des inversions sujet-verbe que nous avions abordé en 1.2.3.2.1 seront compris parmi les pataquès malgré leur caractère normatif. Les avantages méthodologiques à conserver cette définition sont qu'aucune étude préalable n'est nécessaire à l'annotation de liaisons et que cela facilite

[53] En fonction de nos connaissances graphophononologiques en matière de liaison (voir 1.2.3.2.1).

le traitement informatique de nos données. Ce n'est pas forcément la meilleure option, mais il me semble bien avoir pris le temps jusqu'à présent de lister les avantages et les inconvénients de chacune, la discussion restant ouverte, que ce soit pour apporter une appellation ou une définition plus adéquate au phénomène en question.

Mon objectif dans cette sous-partie était de proposer une définition pour le phénomène de pataquès. Après avoir éliminé l'idée de dépeindre les pataquès comme des liaisons non normatives, des liaisons non standard, des liaisons non étymologiques ou encore des liaisons non historiques, j'ai consenti à les caractériser comme des liaisons inattendues, avec un critère graphique. Afin de bien concevoir ce que n'est pas un pataquès, je présenterai dans le point suivant un état de l'art qui recense d'autres productions erratiques en matière de liaison.

1.3.3 Autres productions erratiques en matière de liaison

Il existe des liaisons produites de manière erratique, mais conformes à la consonne graphique finale du M1, que nous avions mentionnées en 1.2.2.3.2. Gadet (1989 : 56) range ces liaisons dans la catégorie des *liaisons par excès*. Celles-ci ne sont pas comprises dans notre catégorie pataquès.

Dans sa thèse, Mallet (2008 : 93) aborde aussi le cas des non-productions de liaisons catégoriques, comme dans *son* [on] *igloo* dans l'exemple 260 et *mes* [oz] *entrailles* en 261 (cf p. 252). En acquisition, on parlera d'***erreur par omission*** (Chevrot et al., 2013 : 87), lorsque l'enfant ne produit pas une liaison qui est invariable chez l'adulte (comme dans *les amis* [leami][54]).

Enfin, Gadet (1989 : 57) présente ce qu'elle nomme des *liaisons par défaut*, qu'elle catégorise comme un cas de « fautes de liaison », pour les « liaisons facultatives fréquentes ostensiblement non faites ».

Pour ce qui est des pataquès, ils se déclinent en deux sous-catégories en acquisition : les ***erreurs par remplacement*** (du type *les amis* [lenami]) et les ***erreurs par adjonction*** (du type *l'ami* [lənami] ; Chevrot et al., 2013 : 87). Nous reviendrons sur celles-ci dans la sous-partie suivante (1.3.4). Mallet (2008 : 96) utilise de façon inappropriée l'appellation de *liaisons erratiques* pour parler des pataquès. Cela porte à confusion car si ceux-ci sont bien erratiques et apparentés aux liaisons, nous avions vu en 1.2.2.3.2 que cette appellation se référait à un type de liaisons en particulier. Je préfère distinguer ces deux phénomènes. Non-réalisations de liaisons invariables et pataquès partageant un

[54]Tous les exemples acquisitionnels sont proposés dans Chevrot et al., 2013 : 87.

caractère erratique, Mallet (2008 : 93) les regroupe dans une catégorie qu'elle nomme *productions agrammaticales*. Il faut y voir ici des réalisations trop erratiques pour relever de la grammaire des locuteurs. En ce qui me concerne, le cadre théorique que j'adopte ne me permet pas d'utiliser une telle dénomination, car comme le dit Goldberg, "the constructionist approach does not assume that language should be divided up into 'core' grammar and [...] 'periphery'" (Goldberg, 2003 : 223). Je ne ferai pas d'autre proposition et me contenterai de parler de productions erratiques (à ne pas confondre avec les liaisons erratiques !) pour l'ensemble de ces phénomènes, ce qui ne pose pas de problème chez l'adulte mais pourrait en poser chez l'enfant (les occurrences en marge du phénomène de liaison sont-elles réellement produites de manière erratique par rapport aux liaisons dans les corpus acquisitionnels ?).
Dans le domaine de l'acquisition, un dernier cas spécifique a été relevé : les ***erreurs par régularisation*** (Chevrot et al., 2013 : 87), lorsque l'enfant remplace la consonne fixe initiale du M2 par une consonne de liaison suite à une métanalyse (comme dans les nombrils [lezɔ̃bʁil][55]).

Un ultime cas de figure semble ne pas avoir été pris en compte : lorsque la consonne fixe initiale du M2 est supprimée suite à une métanalyse où elle a été prise pour une consonne de liaison, hors cas de liaison. C'est ce qu'on a par exemple pour *l'orange* en diachronie (arabe *nārang(a)*), cas de déglutination déjà mentionné en 1.2.3.1.1.3.

En résumé, on observe six types de productions erratiques en matière de liaison :

- des non-réalisations de liaisons invariables (nommées *erreurs par omission* en acquisition),
- des non-réalisations de liaisons variables fréquentes,
- des réalisations de liaisons erratiques (conformes à la graphie du M1),
- des réalisations de pataquès (non conformes à la graphie du M1), appelées *erreurs par remplacement* ou *erreurs par adjonction* en acquisition,
- des non-réalisations de consonnes initiales fixes mais réalisations de liaisons (ou *erreurs par régularisation* en acquisition) et
- des non-réalisations de consonnes initiales fixes prises pour des liaisons, hors cas de liaison.

Rappelons qu'une liaison peut être ou non réalisée, les trois premiers types de productions erratiques seront donc traités lors de l'étude du phénomène de liaison. Les

[55]Rappelons au passage que, comme vu en 1.2.3.1.1.3, la consonne initiale de *nombril* n'est pas étymologique.

productions enfantines ne correspondant pas aux productions adultes relèvent quant à elles du champ de l'acquisition du langage. En un mot, pour clarifier ma position, selon la définition de pataquès donnée en 1.3.2.4, tous les points ci-dessus en dehors de la catégorie pataquès ne peuvent être catégorisés comme tels. Nous allons dans le point suivant spécifier les différents types de pataquès que mentionne la littérature.

1.3.4 Sous-catégories de pataquès

Une première façon de distinguer les pataquès est de regarder dans quel contexte ils sont produits. Nous avions vu dans le point précédent (1.3.3) que dans le domaine de l'acquisition, on distinguait les *erreurs par remplacement*, lorsque ce n'est pas la bonne consonne de liaison qui est produite, des *erreurs par adjonction*, quand une consonne de liaison survient hors contexte de liaison. On pourrait parler dans le premier cas de ***pataquès en contexte de liaison*** (exemple 43) et dans le second de ***pataquès hors contexte de liaison*** (exemple 44). Toujours en acquisition, Liégeois (2014 : 274) parle de la même façon d'*erreurs en contexte de liaison* et d'*erreurs hors contexte de liaison*.

(43) *SGR : dans ma chambrett(e) tu n'as pas bien [*z] été j'ai mon yyy mon honneur et ma fidelité.
%pho : dɑ̃ ma ʃɑ̃bʁɛt ty na pa bjɛ̃ zɛte ʒɛ mɔ nam mɔ nɔnœʁ e ma fidɛlite
(JD_a347_06, 43200, Annette Rousset)

(44) *SGR : ce fut par un beau dimanche qu'ell(e) a dansé [*ɾ] avec [*ə] moi.
(JD_a165_17, 43800, Germaine Varenne)

Une seconde façon de distinguer les pataquès est de proposer une explication linguistique quant à l'origine de la consonne de pataquès. Avec sa compilation de collections d'exemples de seconde main couplée à sa collection d'exemples de première main, Desrochers (1994 : 244) ainsi que Gadet (1989 : 56), toujours sur collection d'exemples, ont remarqué deux types de pataquès. Desrochers (1994 : 261) oppose ce qu'il nomme des *liaisons analogiques* à des *liaisons à distance*. Gadet (1989 : 56) utilise pour sa part les appellations de ***pataquès paradigmatiques*** et de ***pataquès syntagmatiques***. Les premiers semblent liés à des liaisons réalisées dans un autre énoncé sur l'axe paradigmatique : dans l'exemple 44, cet énoncé pourrait être *pour danser* [ʁ] *avec moi*. Les seconds semblent liés à un mot de la chaîne parlée antérieur au M1 : dans l'exemple 43, *pas* pourrait être à l'origine du pataquès en [*z], même si ce dernier est produit après *bien*. Si les premiers pencheraient davantage du côté de la morphologie comme nous le verrons ci-dessous, les seconds seraient d'ordre syntaxique (Desrochers, 1994 : 279).

Pour ce qui est des pataquès paradigmatiques, ils émergeraient d'après Gadet (1989 : 56) « par prégnance de mémoire ». Ainsi, selon elle, ce type de pataquès remobiliserait notre expérience d'auditeur en tant que locuteur. La question est de savoir si une construction lacunaire peut être remobilisée : par exemple, Desrochers (1994 : 248, 261) indique qu'« il [est] souvent possible de faire porter à une analogie reposant sur la catégorie lexicale la responsabilité des fausses liaisons », comme dans le cas des adjectifs prénominaux liant en [t] comme *petit* ou *grand* qui pourraient être à l'origine de *gros* [*t] *enfant*. Lorsqu'une consonne de liaison se fait rare, elle pourrait même être remplacée par une consonne plus fréquente apparaissant dans le contexte en question (par exemple, *trop* [*z] + Adjectif (ADJ) pourrait avoir été construit sur le modèle de très [z] + ADJ ; Desrochers, 1994 : 269). Enfin, d'après Desrochers (1994 : 266), les liaisons analogiques émergeraient hors contexte de liaison invariable.
En ce qui concerne l'appellation de *liaison à distance*, elle apparaît déjà chez Pichon, qui voit dans ce phénomène une persistance de l'image mentale du terme liaisonnant (Pichon, 1935 : 339-340). On pourrait se demander si selon lui cette image mentale comporte une information graphique. Desrochers (1994 : 262) en note peu dans sa compilation de collections d'exemples. Il y voit une « réorganisation du plan de l'énoncé prévu par le locuteur » (Desrochers, 1994 : 279). Les pataquès syntagmatiques posent diverses questions : tout d'abord, il n'y a plus seulement un M1, mais un mot potentiellement liaisonnant non adjacent au pataquès et un mot précédant le pataquès (Desrochers, 1994 : 277). Ensuite, se pose la question de la distance maximale possible entre ces deux mots (Desrochers, 1994 : 268).
Terminons en mettant au jour une différence qui semble exister entre les propositions de Desrochers et de Gadet : la seconde semble considérer les pataquès d'origine morphologique (ne sont mentionnés que les cas ayant trait à la marque de pluriel) comme des pataquès syntagmatiques (Gadet, 1989 : 56). En effet, parfois la valeur morphologique en question semble déjà présente dans l'énoncé (par exemple, « l'idée de pluriel » comme le note Gadet, 1989 : 56). Mais Gadet (1989 : 56) convainc peu puisqu'elle fait elle-même appel au concept de paradigme en parlant des pataquès syntagmatiques. J'adopterai la position de Desrochers (1994 : 277), qui considère les pataquès morphologiques comme des pataquès paradigmatiques.

Ainsi, sont distingués d'une part les pataquès en contexte de liaison des pataquès hors contexte de liaison, selon le lieu d'apparition des consonnes, et d'autre part les pataquès paradigmatiques des pataquès syntagmatiques, selon l'origine présumée de la consonne produite. Nous allons à présent aborder le phénomène de pataquès sous

l'angle linguistique, en commençant par le niveau phonologique.

1.3.5 Le pataquès : facteurs linguistiques

1.3.5.1 Pataquès et phonologie

1.3.5.1.1 Nature et prégnance des consonnes de pataquès

Dans mon mémoire de deuxième année de master, j'avais relevé les consonnes de pataquès suivantes : /z, t, ʁ, n, k/ (Coutanson, 2017 : 73). Voici des exemples issus du FJD :

(45) *SGR : peut + êtr(e) avant troi(s) jours ce sera [*z] à mon tour.
(JD_a160_11, 43800, Jean-Pierre Arnaud)

(46) *SGR : la petite Marianne s'en va [*t] au moulin.
(JD_a161_04, 43800, Jacques Rogues)

(47) *SGR : qui a logé [*r] ici dedans ?
(JD_a165_23, 63400, Valentine Alibert)

(48) *SGR : comm(e) un vrai [*n] artisan.
(JD_a287_04, 63550, Marie Fontenille)

(49) *SGR : je la teindrai de bout [*k] en bout la couleur qui est à votre goût.[56]
(JD_a137_04, 63250, Marie Fayet[57])

Selon la définition du pataquès retenue, il est logique que les consonnes de pataquès observées soient de même nature que les consonnes de liaison. Notons au passage que Desrochers (1994 : 258) remarque lors de sa compilation de collections d'exemples que certains pataquès en /n/ apparaissent après des voyelles nasales, qui pourraient avoir favorisé selon lui le choix de cette consonne.
En outre, dans la littérature, Desrochers (1994 : 245, 247) entre autres observe des pataquès en /p/. Puisqu'il s'agit d'une consonne de liaison, ceci est tout à fait probable. Le seul exemple que j'ai rencontré (exemple 159 p.221), en carottant le fonds Charles Joisten, est discutable car on pourrait y voir une liaison phonologique ayant une réalisation phonétique non prototypique due à une harmonie consonantique (sur le modèle de l'exemple 12 p.30). Morin (1982) mentionne aussi l'existence de pataquès en /l/ en français québecois (voir un exemple en 1.3.6.3). Si les liaisons en /l/ ne sont pas

[56] Il s'agit de l'unique occurrence de pataquès en [*k] que j'avais relevée (toutefois répétée par la chanteuse).

[57] Cette chanson faisait partie de la collection d'exemples utilisée pour mon mémoire de master 2 (Coutanson, 2017) mais est absente des chansons du FJD transcrites et annotées dans le cadre de cette thèse.

systématiques (cf 1.2.2.1.3.1), ce n'est pas pour autant qu'on ne peut pas rencontrer de pataquès en /l/.

Les pataquès en /z/ et /t/ avaient déjà amplement été observés ; ceci se note aisément quand on s'intéresse aux appellations qui existent déjà : comme nous l'avions vu en 1.3.1, les *velours* désignent traditionnellement les pataquès en /z/ et les *cuirs* les pataquès en /t/. J'avais pour ma part proposé, suivant les conseils de Daisy Bonnard, alors en poste au Musée des Tissus de Lyon, l'appellation *moires* pour les pataquès en //ʁ// (Coutanson, 2017 : 74), que j'avais notés de façon significative dans ma collection d'exemples effectuée sur une partie du FJD (env. 10%, cf tableau de la figure 1.4 en 1.2.2.1.3.3).

Dès 1929, Frei, se fondant sur une collection d'exemples, indiquait que les velours étaient plus fréquents que les cuirs (Frei, 1929 : 103-104). Ceci est en accord avec l'ordonnancement des consonnes de pataquès que j'ai observé dans mes deux mémoires de master, sur une collection d'exemples systématique d'une partie du FJD : /z/ > /t/ > //ʁ// > /n/ = /k/ (voir tableau de la figure 1.4[58] en 1.2.2.1.3.3). Cependant, cela va à l'encontre de ce que Hornsby (2012 : 76) a observé en lecture, ce qui pourrait dépendre du matériel d'élicitation utilisé : ce linguiste a remarqué pour une tâche de lecture proposée à des lycéens (présentée en 1.2.4.1) l'ordonnancement suivant : /t/(56) > /z/(23) > /n/ (2)[59].

Pour terminer, dans mon mémoire de master 2 (Coutanson, 2017 : 76-78), je m'étais intéressée aux consonnes de pataquès qui remplaçaient des consonnes de liaison attendues (en contexte de liaison donc). J'avais remarqué, pour ma collection d'exemples portant sur une partie du FJD, que les consonnes de liaison semblaient davantage remplacées par des consonnes plus fréquentes en liaison (55/79) que moins fréquentes en liaison (24/79). Cependant, ne disposant pas de l'ordonnancement des consonnes de liaison dans le fonds, puisqu'il ne s'agissait pas encore d'un travail sur corpus, cette observation serait à vérifier, car j'avais dû utiliser l'ordonnancement des liaisons dans le corpus PFC.

Ainsi, les consonnes de pataquès attestées sont /t, n, z, ʁ, l/ et très certainement /k/ et /p/, même si pour ces dernières, davantage d'exemples seraient souhaitables.

[58] « ce » signifie qu'il s'agit d'observations effectuées sur une collection d'exemples.

[59] En 2020, avec les données supplémentaires des villes de Perpignan et de Mons doublant le nombre d'enquêtés, Hornsby (2020 : 156) observe au total 97 pataquès en [t] et 43 pataquès en [z], ainsi que 60 réalisations autres : [d], [v]... Ainsi l'ordonnancement /t/ > /z/ est confirmé pour son test de lecture bien que le nombre de pataquès en [n] n'ait pas été actualisé.

En phonétique, certains pataquès pourraient être dus à des harmonies consonantiques. Les pataquès en /z/ et /t/ semblent rétrospectivement avoir été les plus notés. L'existence des termes *velours* et *cuirs* et l'absence d'appellations pour les autres consonnes en témoignent. Mes propres recherches sur une partie du FJD avaient mis en évidence l'ordonnancement des consonnes de pataquès suivant : /z/ > /t/ > //ʁ// > /n/ > /k/, ordonnancement qui diffère de ce qui a été noté en lecture à voix haute. Il faudrait encore vérifier que la majorité des pataquès en contexte de liaison concernent bien des consonnes de liaison plus fréquentes en remplaçant d'autres, moins fréquentes. Intéressons-nous à présent aux segments adjacents aux pataquès.

1.3.5.1.2 Pataquès et segments environnants

Dans mon mémoire de master 1, j'avais relevé au cours d'une analyse distributionnelle que les pataquès apparaissaient en grande majorité dans le contexte V#_#V, soit à la frontière entre deux mots, le M1 présentant une finale vocalique et le M2 une initiale vocalique (Coutanson, 2016 : 46). J'avais aussi relevé les contextes ʁ#_#V et V#_#j (Coutanson, 2016 : 46), puis dans mon mémoire de master 2, le contexte l#_#V (Coutanson, 2017 : 71-72). Ainsi, les pataquès semblent pouvoir émerger dans les contextes suivants : $\left\{ \left\{ \begin{matrix} \text{V} \\ \text{ʁ} \\ \text{l} \end{matrix} \right\} \#_\#[\text{-cons}] \right\}$[60].

Les pataquès semblent donc apparaître préférentiellement dans le contexte V#_#V, voire après une liquide ou devant une glissante. Dans le point suivant, nous nous concentrerons sur les pataquès en //ʁ// à la suite de participes passés, un contexte phonétique parfois identique aux infinitifs, après lesquels la marque d'infinitif //ʁ// peut, sous certaines conditions, apparaître en liaison.

1.3.5.1.3 Pataquès en //ʁ// : marque pseudo-morphologique déclenchée par un contexte phonétique similaire

Nous avions vu en 1.2.3.1.1.3 que diachroniquement, des //ʁ// à la finale d'infinitifs avaient cessés d'être prononcés, avant d'être rétablis pour les verbes en -‹ir› et -‹oir›. Nous avions aussi vu en 1.2.2.2.3 qu'actuellement, le //ʁ// pouvait être considéré comme une marque d'infinitif seulement pour les verbes en -‹er› et -‹ier›. Pourtant, comme le notait Monique Léon (1984 : 4),

[60] Le trait distinctif [-cons] englobe les voyelles et glissantes.

> les formes de l'infinitif et du participe passé se recouvrent à l'oral, exemple *il faut travailler aussi*, et, *j'ai travaillé aussi*. D'où c'est faire montre d'une certaine maîtrise de la langue que de savoir faire la liaison avec *R* au point opportun de la chaîne parlée et c'est un moyen de s'assurer un certain prestige.

Dans mon mémoire de master 2 (Coutanson, 2017 : 97-98), j'avais noté un nombre conséquent de pataquès (15) après des participes passés de verbes en -‹er› (ex : *Elle fut blessée* [*r] *au bras*). Mais de façon beaucoup plus inattendue, j'avais relevé une occurrence de pataquès après un participe passé d'un verbe en -‹ir› (*Qui t'as conduit* [*r] *ici ?*), trace d'une prononciation antérieure similaire pour les infinitifs et les participes passés des verbes de ce tiroir verbal. Contrairement aux liaisons en //ʁ// après des infinitifs, les pataquès en //ʁ// après des participes passés n'ont pas de valeur morphologique mais émergent dans un contexte phonétique similaire (ici, après un même mot phonétique).

Ainsi, l'homophonie entre des participes passés et des infinitifs en M1 peut entraîner l'apparition de pataquès en //ʁ// après des participes passés de verbes en -‹er› et -‹ier› (voire en -‹ir›), bien que la valeur morphologique de la consonne soit ici annulée. Après nous être penchés sur la nature et la prégnance des diverses consonnes de pataquès, nous avons rappelé quels segments étaient connus pour appartenir à leur environnement immédiat, avant de nous arrêter plus spécifiquement sur les pataquès en //ʁ// après des participes passés qui semblent pouvoir être expliqués phonétiquement plutôt que morphologiquement. Nous verrons maintenant ce que le niveau morphologique peut apporter à notre compréhension du phénomène.

1.3.5.2 Pataquès et morphologie : morphologisation des consonnes de liaison

Nous avions vu en 1.2.2.2.3 que d'un point de vue fonctionnel, certaines consonnes de liaison s'étaient vu attribuer un rôle morphologique. L'apparition de certains pataquès est souvent expliquée en faisant appel à ces rôles morphologiques :

- le /z/ comme marque de pluriel va être invoqué lorsque cette consonne survient de façon inattendue après un nom composé (ex : *des pommes de terre* [z] *à l'eau* ; Desrochers, 1994 : 251), après des numéraux (ex : *huit* [z] *hommes* ; Desrochers, 1994 : 250 ; Léon, 1984 : 5 ; Bybee, 2005 : 28), etc... Selon Blanche-Benveniste (1997 : 153), les [z] nominaux « ne marquent plus [...] le pluriel d'un mot en

particulier, (on ne comprendrait pas qu'ils puissent s'appliquer à des adverbes ou des conjonctions) mais le pluriel de tout un syntagme ».

- dans mon mémoire de master 2, j'ai invoqué le /z/ comme marque de participant au discours lors de l'utilisation d'un temps composé (*J'ai* [*z] *entendu pleurer* ; Coutanson, 2017 : 175) ou après un pronom enclitique d'un verbe à l'impératif (*Apprends moi* [*z] *à parler !* ; Coutanson, 2017 : 176). J'avais par exemple observé pour les consonnes de pataquès apparaissant avec un temps composé (voir 1.3.5.3) qu'elles étaient généralement en accord, dans 24 cas sur 31, avec le sujet ([*z] quand il s'agissait d'une première ou deuxième personne, [*t] quand il s'agissait d'une troisième personne ; Coutanson, 2017 : 86-87).
- le /t/ comme marque de troisième personne va être invoqué lorsque que cette consonne survient de façon inattendue après un verbe conjugué à la troisième personne (ex : *il y a* [t] *une maison* ; Desrochers, 1994 : 256 ; Léon, 1984 : 5). J'avais pour ma part invoqué cette fonction lors de l'utilisation d'un temps composé (*Sont venus* [*t] *à passer* ; Coutanson, 2017 : 184). Loin l'idée de Blanche-Benveniste (2010 : 53), selon laquelle ce [t] « [...] n'a plus aucun rapport avec la morphologie du mot qui précède [...] », puisque cette consonne est étymologique (voir 1.2.3.1.1.3 et 1.2.3.1.2.2) et paradigmatique (par commutation avec d'autres verbes que ceux en -‹er› ou -‹ier›). Selon Morin et Kaye (1982 : 323), les cas de pataquès postverbaux, exclusivement en /t/ dans leurs observations, indiqueraient une transition d'une marque de personne à une marque verbale. Celle-ci se serait étendue à d'autres contextes dans le groupe verbal : plus seulement après les auxiliaires et les verbes, mais également après des participes passés, l'adverbe de négation *pas* ou encore des enclitiques. Blanche-Benveniste (2010 : 77) fait la même observation : « [certains locuteurs] marquent [un] [t] à toutes les frontières vocaliques du syntagme verbal ». Ces observations ne sont pas valables pour le FJD, qui présentait deux marques verbales, /z/ et /t/.

L'étude des pataquès a permis de proposer de nouveaux rôles morphologiques pour certaines consonnes de liaison/pataquès. Il s'agit notamment du /z/ comme marque d'impératif (ex : *donne-moi*-[z]-*en* ; Gaatone, 1979 : 329 ; Desrochers, 1994 : 254-255). J'avais pour ma part observé un pataquès en [*t] dans ce contexte : *Demandez-le* [*t] *à mon père* (Coutanson, 2017 : 178). Étant donné que les formes impératives en français sont uniquement à la première ou à la deuxième personne, cela expliquerait cette proposition qu'on retrouve dans la littérature. Cela étant, je pense que l'analyse de /z/ comme marque d'impératif est redondante : une analyse générale du /z/ comme marque de participant au discours me semble préférable (économie d'une fonction ; Coutanson,

2017 : 89).

Comme nous l'avions vu en 1.2.2.2.3, la question est de savoir si les locuteurs ont réellement morphologisé des consonnes de liaison, ce que semblent indiquer les cas de pataquès venant d'être mentionnés. Rappelons que l'existence de ceux-ci constitue l'un des deux arguments permettant de pencher en faveur d'une telle morphologisation, avec celui d'une réalisation différentielle des liaisons selon des critères morphologiques. Un avis nuancé avait été émis à propos de ce second aspect, car si cette réalisation différentielle existe bien, la réalisation de consonnes de liaison selon des critères morphologiques n'est pas systématique. Il en va de même pour les pataquès, qui ne sont pas inéluctablement produits dans les contextes susmentionnés pour marquer le pluriel ou encore pour distinguer les personnes d'allocution des troisièmes personnes.

Retenons donc que les pataquès en /z/ sont justifiés en faisant appel au /z/ comme marque de pluriel après des noms composés ou des numéraux, comme marque de participant au discours pour les temps composés ou des verbes à l'impératif, voire comme marque d'impératif après des verbes conjugués à ce mode + enclitique et les pataquès en /t/ en faisant appel au /t/ comme marque verbale, après un verbe, un participe passé, l'adverbe de négation *pas* ou des enclitiques. Si la production de pataquès va bien dans le sens d'une morphologisation des consonnes de liaison, à nouveau, comme pour la réalisation différentielle des consonnes de liaison selon des critères morphologiques, les occurrences produites ne sont pas assez systématiques pour que la morphologisation ait été complète. Nous adopterons à présent un angle d'approche syntaxique, en recensant notamment les contextes morphosyntaxiques mentionnés dans la littérature comme des contextes potentiels d'apparition de pataquès.

1.3.5.3 Pataquès et syntaxe : contextes syntaxiques d'apparition de pataquès connus

La présentation qui suivra des contextes morphosyntaxiques dans lesquels des pataquès émergent parfois découle principalement des observations de Desrochers (1994), fondées sur une compilation de collections d'exemples[61]. Les contextes singuliers sont indiqués à l'aide de la glose SG, les contextes pluriels à l'aide de la glose PL. Des pataquès ont été notés dans les contextes suivants :

[61] Certaines catégories ont été modifiées et certains exemples n'ont pas été retenus, notamment les pataquès en cas de géminées. Les abréviations utilisées sont principalement celles du jeu d'étiquettes de Treetagger pour le français ; cf Stein, 2003.

- ADJ.SG + NOM.SG (*gros* [t] *écran*, *un grand* [n] *orgueilleux* ; Desrochers, 1994 : 246)
- ADJ/NOM/Numéral (NUM) + Conjonction (KON) (*Où allez-vous jolie* [*z] *et belle ?* ; *Mon père* [*z] *et ma mère* ; *Des mille* [*z] *et aussi des millions* ; Coutanson, 2017 : 171, 185, 186)
- Interne à une locution adverbiale (*peu* [z] *à peu* ; Desrochers, 1994 : 246)
- Adverbe (ADV) de degré + ADJ.PL (*Mes parents seraient trop* [z] *heureux* ; Desrochers, 1994 : 246), ou plus généralement, ADV de degré + ADJ (*t'es trop* [z] *indiscret* ; Desrochers, 1994 : 246)
- ADV + ADV (*nous pouvons penser que les choses se passent de même* [t] *ici* ; Pichon, 1935 : 340)
- ADV + NOM (*je n'en ai aucunement* [z] *honte* ; Desrochers, 1994 : 246)
- ADV + Syntagme prépositionnel (SP) (*c'est pas* [t] *à moi* ; Morin & Kaye, 1982 : 324)
- ADV + VER (*déjà* [z] *allé*, *peu* [n] *importe* ; Desrochers, 1994 : 246 ; *ils ne veulent pas* [t] *aller à l'école* ; Morin & Kaye, 1982 : 324)
- Préfixe + NOM.PL/ADJ.PL (*les non*-[z]-*handicapés*, *ces micro*-[z]-*histoires* , *des organisations pro*-[z]-*iraniennes* ; Blanche-Benveniste, 2010 : 57, voir aussi Desrochers, 1994 : 251)
- Locution nominale.PL + ADJ.PL/NOM.PL/SP (*chemins de fer* [z] *algériens*, *pour trouver tous les valets de chambre* [z] *horlogers*, *des pommes de terre* [z] *à l'eau* ; Pichon, 1935 : 339)
- NOM + ADV (*pour vous présenter mes idées* [r] *ici* ; Pichon, 1935 : 339)
- NUM.PL + NOM.PL (*quat'* [z] *enfants* ; Frei, 1929 : 98 ; *plein de* [z] *habits* ; Pichon, 1935 : 341)
- Préposition (PRP) + Déterminant (DET)(Syntagme nominal (SN)) (*devant* [z] *un* ; Desrochers, 1994 : 246)
- PRP + NOM.PL/Pronom (PRO).PL (*c'est quoi comme* [z] *arbres* ; Morin, 1986 : 195 ; *parmi* [z] *elles* ; Frei, 1929 : 104), ou plus généralement PRP + PRO (*bien malgré* [n] *elle* ; Desrochers, 1994 : 246)
- PRO + ADV (*moi* [z] *aussi* ; Frei, 1929 : 103)
- PRO + KON (*moi* [z] *et lui* ; Frei, 1929 : 103)
- (Verbe conjugué à l'impératif présent (VER:impe) +)PRO + PRO/ADV/SN (*menez-moi* [z] *y*, *prends-moi çte brique et fous-lui* [z] *y sur la gueule* ; Frei, 1929 : 103 ; *laisse-moi* [t] *ici*, *donne-moi* [t] *un conseil* ; Morin & Kaye, 1982 : 324)
- PRO personnel objet *leur* + VER (*Les feuilles que je leur* [z] *ai données* ; Desrochers, 1994 : 246), sûrement par analogie avec le DET possessif de 3.PL *leurs*

- Pronom indéfini (PRO:IND) *tous* + SP (*ils chantent tous* [t] *en cœur* ; Morin & Kaye, 1982 : 324)
- Complémenteur[62] *qui* + VER.3.PL (*les enfants qui* [z] *arrivent* ; Gadet, 1992 : 49, voir aussi Desrochers, 1994 : 251). Gadet (1995 : 142) considère ces cas comme des *relatives non standard*. Selon la linguiste, il s'agirait non de *qui* mais de *qu'i(l)s*, une « structure qui n'est révélée que par la liaison, donc au pluriel, puisqu'il est courant de ne pas prononcer le [l] de il(s) devant consonne ; elle est plus rare au singulier », même si on trouve des exemples tels que *C'est lui qu'il arrive en général au travail à 10 heures* (Gadet, 1995 : 142). Ces exemples viennent confirmer qu'une telle structure se rencontre et que la présence du pronom n'est pas que théorique et hypothétique. Abeillé et Godard (2007 : 40) parlent de *relatives en* qui *sans gap*, puisque contrairement aux relatives standard (*avec gap*), elles présentent un pronom. Bien qu'il ne faille pas tomber dans le piège de chercher à tout prix une explication d'un fait linguistique non standard à partir du standard, on pourrait néanmoins supposer que ce phénomène trouve son origine dans l'existence d'un conflit entre deux constructions : *qu'ils* + VER.3.PL vs *qui* + VER.3.PL, qui peuvent se réaliser de la même façon devant initiale consonantique (exemple 50) mais différent systématiquement devant initiale vocalique (exemple 51).

(50) *Ce sont les enfants qui m'écoutent* [kimekut] vs *Je veux qu'ils m'écoutent* [kimekut]/[kilmekut]

(51) *Ce sont les enfants qui arrivent* [kiaʁiv] vs *Je veux qu'ils arrivent* [kizaʁiv]/[kilzaʁiv]

Au singulier, Gadet (1995 : 142) inclut en outre les exemples comportant une géminée tels que *C'est moi qui ai dit c'est celui-là que je veux mais c'est eux* [killɔ̃] *payé*. Les géminées latines ont été simplifiées au VII^{ème} s. (Laborderie, 1994). Les seules géminées qu'on rencontre en français contemporain résultent de rencontres entre segments de même nature dans la chaîne parlée, comme *à t(ou)t à l('heure)* [attal(œʁ)], ou d'orthographismes pour des mots comprenant des consonnes doubles tels que *collègue*, *illustre*, *inné* ou encore *sommet* (Russo & Sahmaoui, 2021 ; Brandão de Carvalho, 2018 : 48). L'analyse des réalisations comme [killɔ̃] sur le modèle de [kizɔ̃], avec latence d'un pronom sujet de 3.SG/PL masculin ou par analogie avec une autre construction (*qui* + *l'* + VER.3.SG vs *qu'il* + *l'* + VER.3.SG ou *qui* + VER.3.PL vs *qu'ils* + VER.3.PL) ne permet toutefois pas d'expliquer des occurrences telles que *tiens j(e) te* [*l] *l'ai ram(e)né* (exemple p. 218, issu du carottage

[62] Pour savoir pourquoi *qui* est considéré ici comme un complémenteur, cf Abeillé & Godard, 2007 : 38-39.

du corpus ALIPE). On peut supposer que ces dernières proviennent de réalisations géminées comme *il l'a dit* [illadi] ou *elle l'a dit* [ɛlladi], la consonne du pronom sujet de 3.SG rencontrant la consonne d'un pronom objet de 3.SG élidé, ce dernier étant par la suite réanalysé comme [ll] (Brandão de Carvalho, 2018 : 49), avant de se répandre par analogie après d'autres pronoms sujets ou objets. Morin (2019 : 73-74) présente un état de l'art des différentes hypothèses qui ont été formulées pour expliquer cette gémination du pronom objet élidé de 3.SG *l'*, état de l'art qui ne se limite pas aux trois hypothèses déjà présentées. Ainsi, si on analyse systématiquement les relatives susmentionnées comme des relatives non standard, on ne considérera pas ces productions linguistiques comme des pataquès. Au contraire, si on émet le doute que toutes les occurrences actuelles soient des relatives non standard, sans nier en bloc leur existence, on pourra admettre que certaines réalisations constituent des pataquès. Ce sera ma position, renforcée par l'existence de géminées hors relatives après d'autres pronoms que les pronoms sujets de 3.SG et de 3.PL. Cette position n'est pas définitive, mais je n'ai pas le temps de m'attarder outre mesure sur ces réalisations dans le cadre de cette thèse.

- Auxiliaire + Participe passé (*grâce aux maîtres que j'ai* [z] *eus* ; Frei, 1929 : 104). Dans mon mémoire de master 2 (Coutanson, 2017 : 84-86), j'avais noté qu'un pataquès pouvait être produit, dans ma collection d'exemples fondée sur des chansons de tradition orale, soit entre l'auxiliaire et le participe passé (5/20), soit après le participe passé (14/20), soit après l'auxiliaire et après le participe passé (1/20). Dans le premier cas, tous les mots suivants les participes passés étaient à initiale consonantique (*J'ai* [z] *entendu pleurer* ; Coutanson, 2017 : 175), alors que dans le second cas, tous les mots suivants les participes passés étaient à initiale vocalique (ex : *Je ne suis pas venu* [z] *ici pour boire* ; Coutanson, 2017 : 188). On pourrait donc en déduire, même s'il est nécessaire de rester extrêmement prudents vu le faible nombre d'occurrences, que lorsqu'un pataquès apparaît avec un temps composé dans ce corpus, il est produit de préférence après le participe passé, sauf si le mot suivant est à initiale consonantique, auquel cas il est produit entre l'auxiliaire et le participe passé (il est alors nécessaire que ce dernier soit à initiale vocalique). Dans un seul cas, lorsqu'à la fois le participe passé et le mot le suivant étaient à initiale vocalique, le pataquès a été produit après l'auxiliaire et après le participe passé (ex : *J'ai* [t] *aperçu* [t] *une bergère* ; Coutanson, 2017 : 191), mais ce double marquage n'était pas systématique, comme le montrait un contre-exemple (*J'ai entendu* [t] *une jeune voix* ; Coutanson, 2017 : 190).
- (Auxiliaire) + Participe passé + ADV (*je suis né* [z] *ici* ; Desrochers, 1994 : 246)

- VER + ADJ.PL (*mes parents étaient* [z] *italiens* ; Desrochers, 1994 : 246)
- VER + DET(SN) (*dessous Paris y a* [t] *un grand bois* ; Coutanson, 2017 : 169)
- VER + KON (*il va* [t] *et vient* ; Frei, 1929 : 104)
- VER + PRO/VER (*il devra* [t] *y aller* ; Morin & Kaye, 1982 : 297 ; *c'est d'avoir quelque chose qui puisse* [t] *être très général* ; Desrochers, 1994 : 246)
- VER + SP (*i(l) va* [t] *en ville* ; Frei, 1929 : 104 ; *je vais retourner* [z] *à mes quartiers* ; Desrochers, 1994 : 246)

Pourtant, comme pour les liaisons, les pataquès sont très inégalement produits dans ces différents contextes morphosyntaxiques, et il est nécessaire de mettre au jour les contextes préférentiels d'apparition de pataquès. D'après Desrochers (1994 : 252, 259), ce sont NUM + NOM, PRP + PRO, ADV de degré + ADJ, VER + VER et ADJ + NOM.

Comme on le voit, la littérature regorge d'exemples de pataquès, qui permettent de relever quels contextes morphosyntaxiques semblent productifs. Ces données sont issues de collections d'exemples, non attestés (à l'exception de mes exemples de master) : il est ainsi permis de douter de leur réalisation, et de ce fait, des contextes retenus dans cette sous-partie. Ceci serait de la sorte à confirmer à l'aide de données attestées, qu'il s'agisse de collections d'exemples situés ou de données de corpus. La productivité des différents contextes est variable. Enfin, j'ai également proposé une réflexion sur l'analyse à avoir des relatives en *qui* + [z] + VER.3.PL ou *qui*/PRO sujet(+ PRO objet) + [l] + PRO objet *l'* + VER. Nous aborderons à présent notre phénomène d'étude sous un angle sémantique.

1.3.5.4 Pataquès et sémantique

Un calembour peut apparaître lors de la réalisation d'un pataquès : j'avais mentionné dans mon mémoire de master 1 le vers suivant, issu d'une collection d'exemples de chansons de l'industrie du disque (Coutanson, 2016 : 55-56) :

(52) Je veux [*r] aller où le soleil brille (ce, Eddy Mitchell, *Comme un étranger dans la ville*, 2009)

Ici, Eddy Mitchell produit un pataquès en [*r], et on peut comprendre *Je veux râler où le soleil brille* plutôt que *Je veux aller où le soleil brille*. Le second sens était pourtant celui qu'on retrouvait dans la chanson originale :

(53) Je veux [z] aller où le soleil brille (ce, Gilles Marchal, *Comme un étranger dans la ville*, 1969)

Ainsi, la réalisation d'un pataquès peut parfois entraîner un changement de sens et mener au calembour.

Nous avons abordé le phénomène de pataquès sous un premier angle d'approche linguistique, tout d'abord au niveau phonologique (et phonétique), puis au niveau morphologique, au niveau syntaxique et enfin au niveau sémantique. Tournons-nous dès lors vers les facteurs sociolinguistiques qui pourraient expliquer l'apparition de pataquès, en commençant par la dimension diachronique.

1.3.6 Le pataquès : facteurs sociolinguistiques

1.3.6.1 Pataquès et variation diachronique

Le pataquès reste un phénomène encore très peu étudié au niveau diachronique. Nous ne savons pas quand le phénomène est apparu, même si on peut supposer que son avènement est postérieur à l'émergence du phénomène de liaison, si certaines consonnes de pataquès ont commencé à être réalisées plus tardivement ou encore si chaque contexte morphosyntaxique d'apparition de pataquès a été productif à toutes les époques.

Dans mon mémoire de master 2, je m'étais intéressée aux chansons qui avaient plusieurs versions, afin de voir si certains pataquès pouvaient trouver leur origine dans les paroles d'une autre version de la même chanson. Par exemple, j'avais observé les doublons suivants :

(54) Quand l'p'tit bonhomme s'en va-t-au bois – Trouve la tête de son âne – Que le loup avait mangée [*t] au bois (Coutanson, 2017 : 92)

(55) Petit bonhomme s'en va-t-au bois – Trouva l'oreille de son âne – Que le loup mangeait [t] au bois (Coutanson, 2017 : 92)

(56) Il y a bien sept ans [*t] aujourd'hui (Coutanson, 2017 : 92)

(57) Il y a [*t] aujourd'hui sept ans (Coutanson, 2017 : 92)

On voit que les exemples 54 et 55 sont deux versions d'une même chanson, tout comme les exemples 56 et 57, chaque fois interprétées par des chanteurs différents. Dans le premier couple, le pataquès de la première version correspond à une liaison dans la seconde version. J'avais émis l'hypothèse que certaines liaisons avaient pu être conservées sous forme de chunks, soit issus d'une autre version antérieurement interprétée par le chanteur, soit issus d'une autre version à laquelle le chanteur a été exposé. Dans

le second couple, nous faisons face à deux pataquès. Ici, l'hypothèse est qu'un pataquès ayant une explication morphologique (/t/ comme marque de troisième personne) a pu être conservé sous forme de chunk et réutilisé après un M1 pour lequel l'explication morphologique ne fonctionne pas (ici, une marque de troisième personne est incongrue après le mot *ans*), dans une version postérieurement interprétée par le chanteur ou dans la version d'un autre chanteur exposé au pataquès d'origine. Toutefois, les exemples 54 et 56 pourraient aussi être considérés comme des pataquès syntagmatiques.

Mais cela ne s'arrêtait pas là : les chunks semblaient pouvoir être issus de passages différents soit de la même chanson (exemple 58), le pataquès pouvant être antérieur ou postérieur à la liaison ou à l'autre pataquès, soit de versions différentes de la même chanson (exemples 59/60 et 61/62). On remarque que les chunks sont à granularité variable, avec un contexte lexical précis ou uniquement morphosyntaxique mais aussi à taille variable, en incluant le M1 et le M2, seulement le M2, voire en présentant un contexte plus large : par exemple en 58, on observe le schéma SN + [(*)z] + *à*, en 59 et 60, le schéma [(*)z] + *enfants* + *aimables* et en 61 et 62, le schéma PRO personnel sujet + *s'en aller*.PRS + [*t] + SP *à la caserne*.

(58) Tu mettras la main [*z] à la plume – Pour la première fois – Tu mettras les pieds [z] à terre – Pour venir auprès de moi (Coutanson, 2017 : 93)

(59) Oh dites-moi [*z] enfants aimables – Oh dites-moi votre surnom (Coutanson, 2017 : 93)

(60) Où sont nos deux [z] enfants aimables – Mon cher époux tu les verras (Coutanson, 2017 : 93)

(61) Elle s'en va [*t] à la caserne […] (Coutanson, 2017 : 93)

(62) J'm'en vais [*t] à la caserne (Coutanson, 2017 : 93)

Dans mon mémoire de master 2 (Coutanson, 2017 : 110), j'ai également pu étudier la transmission diachronique intra- et inter-chanteurs des pataquès. Dans le premier cas, j'ai observé si ces derniers étaient réitérés au cours d'une même chanson ou dans les performances ultérieures du chanteur, et dans le second cas, j'ai observé s'ils étaient réitérés dans les reprises d'autres chanteurs. En ce qui concerne la transmission intra-locuteurs au sein d'une même chanson, pour le FJD, la plupart des pataquès étaient réitérés (53/59), avec entre une et neuf reprises. Rarement, un pataquès émergeait lors de la reprise, pourtant absent à la première occurrence (4/59, dont 2/59 en contexte de liaison et 2/59 hors contexte de liaison) ou la non-liaison était préférée au pataquès (2/59). En ce qui concerne la transmission intra-locuteurs dans les performances ulté-

rieures de chanteurs de l'industrie du disque, mon analyse se fondait sur des archives audio et télévisées de l'INA. J'avais noté trois stratégies selon les chanteurs ou les chansons concernés : conservation du pataquès (5/10), conservation puis abandon (1/10) ou abandon (4/10) apparents dans la suite de la carrière (l'abandon se faisant alors au profit d'une non-liaison). Si les chanteurs traditionnels semblaient peu en insécurité linguistique en ne corrigeant généralement pas leurs pataquès au sein d'une même chanson, la stratégie d'abandon des pataquès pour certaines chansons de chanteurs de l'industrie du disque semblait dénoter une forme d'insécurité linguistique, quand le maintien du pataquès pourrait découler d'une volonté de fidélité à la performance d'origine. J'ai encore utilisé ces archives de l'INA pour étudier la transmission inter-locuteurs des pataquès, où toutes les stratégies déjà signalées se laissaient apercevoir : abandon du pataquès pour une non-liaison (18/26 reprises), remplacement par la consonne de liaison attendue (2/26), remplacement par un autre pataquès (1/26) et conservation du pataquès (5/26). On voit que l'abandon était la stratégie adoptée préférentiellement par ces chanteurs.

Ainsi, de nombreux questionnements subsistent concernant le phénomène de pataquès en diachronie. J'avais émis par le passé l'hypothèse que certaines réalisations qu'on note dans les chants de tradition orale pourraient provenir d'un chunk issu d'un autre vers de la même chanson ou d'une autre version de la chanson, contenant une liaison ou un pataquès motivé morphologiquement. J'avais également étudié la transmission intra- et inter-chanteurs des pataquès, où j'avais remarqué que les chanteurs traditionnels se corrigeaient rarement, contrairement à certains chanteurs de l'industrie du disque. Nous poursuivrons notre étude du pataquès en nous tournant vers la dimension diamésique, en nous intéressant tout particulièrement aux pataquès écrits.

1.3.6.2 Pataquès et variation diamésique

1.3.6.2.1 Pataquès et graphie

Nous avions vu en 1.2.3.2.1 que la liaison était un phénomène principalement oral et que si certaines consonnes graphiques finales pouvaient donner des indications quant à la nature d'une potentielle consonne de liaison, rien ne laissait présager qu'à l'oral elles seraient réalisées, en-dehors de quelques graphies non standard. Il est encore plus difficile de savoir où les pataquès pourraient émerger (même si le contexte V#_#V est privilégié et que certains contextes grammaticaux sont connus pour être des contextes préférentiels d'apparition) ou encore de quelle nature seraient les consonnes de pata-

quès en question (cette fois-ci aucun indice graphique n'est là pour nous renseigner).

À nouveau, quelques graphies non standard sont parfois utilisées : dans un roman cherchant à mieux représenter l'oral (63), dans la correspondance de peu-lettrés (64 et 65 ; voir 1.2.3.2.1 pour une présentation du corpus)…

(63) – J'sus t'été vérifier qu'un des deux gonziers portait bien la ch'valière qu'on avait causé. […] (Frédérique Dard, San Antonio n°97 : *Si ma tante en avait*, chapitre treize, 2010)

(64) *sont arives ta lamar tunique* (1793, Louis Laniez, Point-à-Pitre, à sa femme au Havre)

(65) *me fait douter que tu mas tecris* (1793, Louis Laniez, Point-à-Pitre, à sa femme au Havre)

Dans mon mémoire de master 1 (Coutanson, 2016 : 70-71, 103-108), j'avais étudié une collection d'exemples restreinte (53 occurrences, parfois bissées) fondée sur des paroles écrites de chansons traditionnelles, issues de recueils ou publiées en ligne. Je m'étais intéressée à la ponctuation utilisée pour noter les pataquès dans ces chansons. J'avais relevé l'ordonnancement suivant : ‹M1 x'M2›[63] (35/58 ; 60,3%) > ‹M1-x-M2› (17/58 ; 29,3%) > ‹M1 x-M2› (3/58 ; 5,2%) = ‹M1-x'M2› (2/58 ; 3,4%) = ‹M1 ʿx' M2› (1/58 ; 1,7%). Cette collection reste bien sûre trop succincte pour être exhaustive et pour que ces résultats soient significatifs, mais cela peut permettre de se faire une première idée. Certaines notations existent ailleurs dans la langue, comme ‹M1 x'M2› qui sert à noter de nombreuses élisions (ex : *j'aime*), ‹M1-x-M2› qui sert à noter les ‹t› dits *euphoniques* lors des inversions sujet-verbe (exemple 32 p.80), ou encore ‹M1-x'M2›, qui sert à noter certaines élisions pronominales en contexte verbal (ex : *va-t'en !*), et pourraient porter à confusion. J'avais constaté que dans un même recueil, on pouvait trouver de la variation. Par ailleurs, si toutes ces utilisations de marques de ponctuation me semblaient dénoter une certaine conscience linguistique de la part des scripteurs quant au rattachement des pataquès au M2 (ressyllabation), les graphies avec un signe de ponctuation présent également avant le pataquès paraissaient davantage les lier au M1 comme au M2. Pour terminer, j'avais observé dans cette collection d'exemples davantage de pataquès en [z] notés ‹ z'› (24/33 ; 72,7%) que ‹-z-› (8/33 ; 24,2%), ce qui laissait entendre que la première graphie pourrait être préférée à la seconde, mais autant de pataquès en [t] notés ‹ t'› (11/25 ; 44%) que ‹-t-› (9/25 ; 36%).
Au-delà de ces graphies relevées dans une collection d'exemples de paroles écrites de

[63] Avec x = toute consonne de pataquès.

chansons traditionnelles, on note dans la correspondance de peu lettrés (exemples 64 et 65) que la consonne de pataquès est accolée au M2, selon le schéma ‹M1 xM2›.
Pour terminer, les linguistes ont parfois opté pour d'autres graphies. Frei note les pataquès selon le schéma ‹M1'*x*[64]M2› (Huit'*z*[64]hommes ; Frei, 1929 : 98), selon le schéma ‹M1 *x*[64]M2› (J'ai chez moi une dizaine de *z*[64]ouvrières ; Frei, 1929 : 98) ou selon le schéma ‹M1-*x*[64]-M2›, également utilisé par P. Léon (Donne-lui-*z*[64]-en ; Frei, 1929 : 103, « Je l'ai-*z*-eu » ; Léon, 1992 : 245), Gadet selon le schéma ‹M1 x M2› (ex : moi z aussi ; Gadet, 1989 : 56) ou ‹M1 x'M2› (ex : donne moi z'en ; Gadet, 1992 : 49), Desrochers et Armstrong & Pooley selon le schéma ‹M1 [x] M2› (ex : *Moins [t] exigeant* ; Desrochers, 1994 : 246, *trop* [z] *âgés* ; Armstrong & Pooley, 2010 : 115), Muller et Blanche-Benveniste selon le schéma ‹M1-x-M2› (ex : parmi-z-eux ; Müller, 1985 : 174, les non-z-handicapés ; Blanche-Benveniste, 1997 : 30) ou selon le schéma ‹M1-x[65]-M2› ou ‹M1-x[65]'M2› pour Blanche-Benveniste (ex : *des portions presque-z[65]-entières* ; Blanche-Benveniste, 2010 : 17, *donne-moi-z[65]'en* ; Blanche-Benveniste, 2010 : 117), P. Léon selon le schéma ‹M1 X-M2› (ex : « Je suis T-allé » ; Léon, 1993 : 201) ou selon le schéma ‹*M1-X-M2*› (ex : « *Je suis-T-allé* » ; Léon, 1992 : 247), M. Léon selon le schéma ‹M1-X-M2› (ex : « mille-Z-îles » ; Léon, 1984 : 5) et enfin Nardy et al. selon le schéma ‹*M1 (**x**) M2*› (ex : *Une puce, un pou jouaient (**z**) aux cartes* ; Nardy et al., 2014 : 255).

Pour conclure, à l'exception de la frontière de mots, il n'existe pas d'indice graphique suggérant la présence de pataquès. Des graphies non standard peuvent à l'occasion être employées pour indiquer cette présence. Dans ce cas-là, j'ai relevé des usages de la ponctuation multiples, dans des contextes très divers (chansons traditionnelles, littérature scientifique, graphies non standard) :

- ‹M1 xM2›
- ‹M1 *x*M2›
- ‹M1 x'M2›
- ‹M1 x-M2›
- ‹M1 X-M2›
- ‹M1-x'M2›
- ‹M1-X'M2›
- ‹M1-x-M2›
- ‹M1-*x*-M2›

[64] Italique
[65] Petite majuscule

- ‹M1-x-M2›
- ‹M1-X-M2›
- ‹*M1-X-M2*›
- ‹M1’*x*M2›
- ‹M1 ʿx’ M2›
- ‹*M1 (x) M2*›
- ‹M1 [x] M2›
- ‹M1 x M2›

Nous nous intéresserons maintenant au pataquès en diatopie.

1.3.6.3 Pataquès et variation diatopique

Que sait-on déjà à propos de la réalisation du pataquès en diatopie ? Dans la littérature, il est fait mention (voir Côté, 2017 : 21-22) :

- de pataquès en [z] :
  - dans la construction *de* + NOM pluriel en Louisiane (*combien d’*[z]*enfants*, *assez d’*[z]*années* ; Boutin & Lyche, 2014 : 298)
  - dans la construction NUM + NOM pluriel dans les variétés américaines (Canada : *vingt* [z] *enfants* de l’exemple 111, Louisiane : *cinq* [z] *heures* ; Boutin & Lyche, 2014 : 298) et européennes (*cent* [z] *euros* de l’exemple 112). Ces pataquès seraient plus fréquents après les nombres à finale vocalique en isolation qu’après les nombres à finale consonantique en isolation (Côté, 2017 : 22). Le français louisianais distinguerait enfin valeur ordinale, sans pataquès, de valeur numérale, avec pataquès potentiel (*il est cinq heures* vs *j’ai fait ça en cinq* [z] *heures de temps* ; Boutin & Lyche, 2014 : 298).
  - dans la construction enclitique + enclitique au Canada (*parle-moi*-[z]*en* ; Côté, 2017 : 22) et en France (voir notamment Blanche-Benveniste, 2010 : 117) :

(66) *MOT : tiens occupe toi [z] en.
(ALIPE, Clermont-Ferrand (France), 2008, ali-prune-080715-1)

  - lexicalisés par certains locuteurs en Louisiane (suite à des agglutinations, voir 1.2.3.1.1.3). On trouve dans le *Dictionary of Louisiana French*[66] : *z-acre/acre* ;

[66] La notation avec un tiret est une convention du *Dictionary of Louisiana French* ne servant qu’à rappeler plus rapidement au lecteur l’équivalent en français standard. Il s’agit d’une véritable lexicalisation et non d’un même lexème prenant différents préfixes selon le contexte. Il n’est pas toujours possible d’après les exemples fournis dans le dictionnaire de savoir si ces pataquès lexicalisés apparaissent hors contextes de

z-agrafe/agrafe ; *z-aignée, z-araignée/araignée* ; *z-ail/ail* ; *z-aile/aile* ; *z-ampoule/ampoule* ; *z-ananas/ananas* ; *z-arbre/arbre* ; *z-haricot, zarico/haricot* ; *z-automobile/automobile* ; *z-enfants* (interjection) ; *z-habitant/habitant* ; *z-hameçon/hameçon* ; *z-herbe/herbe* ; *z-hibou/hibou* ; *z-huître/huître* ; *z-huîtreur/huîtreur* ; *zyeu, zieu/oeil* ; *z-indien/indien* ; *z-intestin/intestin* ; *z-œufs/œufs* ; *z-oie, zoie/oie* ; *z-oignon/oignon* ; *z-oiseau/oiseau* ; *z-oreille/oreille* ; *z-orteil/orteil* ; *z-ouïe/ouïe* ; *z-outil/outil* ; *z-yeux/ yeux* (Valdman et al., 2010 : 11, 17, 18, 27, 35, 48, 326, 329, 332, 333, 336, 343, 346, 424, 425, 428, 429, 431, 664). Boutin et Lyche (2014 : 299) indiquent qu'« il s'agit d'entrées fréquemment utilisées au pluriel », mais comme la liste ci-dessus le montre, les termes à [z] initiaux ne sont pas systématiquement réalisés par tous les locuteurs.

- des pataquès en [t] dans ces constructions-là :
  - X + *être* dans les variétés américaines (Canada : *ça va* [t] *être des nouveautés* : Côté, 2012 : 267, Louisiane : *sa mame va* [t] *être fâchée* ; Boutin & Lyche, 2014 : 300). Ces pataquès ne se produisent pas seulement après des semi-auxiliaires (Côté 2005 : 73 présente les exemples suivants : *j'ai failli* [t] *être, j'voudrais pas* [t] *être, tu peux ben* [t] *être,* qui demeurent en contexte verbal)
  - *suis/es* + X dans les variétés canadiennes (*je suis* [t] *un peu perdu, tu es* [t] *arrivé* ; Côté, 2017 : 22), sachant que le reste du système fait liaison en [t] ou présente la consonne fixe [t] à la frontière entre M1 et M2 (*il/elle/on est* [t], *on est* [t] pour la 1.PL, *vous êtes* ([t] consonne fixe du M1) et *ils/elles sont* [t] ; voir Desrochers, 1994 : 263).
- des pataquès en [l] dans la construction *ça* + X dans les variétés canadiennes (*si ça* [l] *a déjà été vrai* ; voir 1.2.3.1.1.2 ; Côté, 2017 ; Morin, 1982)
- des pataquès en [n] lexicalisés par certains locuteurs en Louisiane. On trouve dans le *Dictionary of Louisiana French* : *n-anglais/anglais* ; *ninorant/ignorant* ; *n-oncle, noncle, nonc/oncle* (Valdman et al., 2010 : 28, 339, 426).

Par ailleurs, dans mon mémoire de master 2 (Coutanson, 2017 : 104-110), j'avais étudié la répartition spatiale de quelques pataquès produits dans certaines chansons de tradition orale par des chanteurs distincts. J'avais distingué trois catégories :

pataquès potentiels en français louisianais, comme dans les contextes *de* + NOM ou NUM + NOM, ce qui pourrait concerner tout mot à initiale vocalique. Si des exemples sont fournis hors de ces contextes, je présente les termes en couleur. Il faut faire preuve de prudence, car rien n'indique que les données soient toutes issues de corpus spontanés ; nous pourrions avoir affaire à des données élicitées. Enfin, les termes peuvent différer dans leur sémantique des termes en français standard.

- les pataquès résultant d'une initiative individuelle, produits par un seul chanteur. Il semblait s'agir des pataquès les moins transparents morphologiquement.
- les pataquès partiellement répandus et transmis localement, produits par plusieurs, mais pas la totalité des chanteurs. Malgré le faible nombre de données, la proximité de résidence de certains chanteurs paraissait pouvoir expliquer certaines (non-)réalisations communes. Bien sûr, pour cela, il était nécessaire que les chanteurs aient été amenés à s'écouter chanter les uns les autres, afin qu'une réelle influence ait pu avoir lieu.
- les pataquès largement répandus et transmis localement, produits par la majorité, voire l'ensemble des chanteurs.

Si ces observations sont faites en synchronie, en diachronie les pataquès ont pu se répandre à davantage de chanteurs ou au contraire se cantonner à un cercle de chanteurs plus restreint.

Récapitulons : en diatopie, ont été mentionnés : des pataquès en [z] dans les constructions *de* + NOM pluriel en Louisiane, NUM + NOM pluriel dans les variétés américaines et en France, enclitique + enclitique dans les variétés canadiennes et en France, dans les noms ou adjectifs précédés d'un préfixe en France, dans la relative en *qui* + VER conjugué à la 3.PL en France ainsi qu'à l'initiale de quelques termes lexicalisés en français louisianais ; des pataquès en [t] dans les constructions X + *être* dans les variétés américaines et *suis/es* + X dans les variétés canadiennes ; des pataquès en [l] dans la contruction ça + X dans les variétés canadiennes et des pataquès en [n] à l'initiale de quelques termes lexicalisés en français louisianais. J'ai aussi émis l'hypothèse qu'une transmission ou perte aréale de pataquès puisse s'être produite pour les chansons de tradition orale. Nous terminerons en nous intéressant au pataquès en **diagénie**, qui concerne la variation que l'on observe dans la langue selon le sexe du locuteur, de l'interlocuteur ou encore du référent.

1.3.6.4 Pataquès et variation diagénique

Lors d'une étude où des lycéens se sont vus proposer une tâche de lecture, présentée en 1.2.4.1, Hornsby (2020 : 157) a constaté davantage de pataquès chez les adolescentes (125/196) que chez leurs homologues masculins (71/196). Par ailleurs, cet auteur avait auparavant remarqué que sur 26 réparations de pataquès faites par des filles, treize l'ont été au profit de la bonne consonne de liaison, douze au profit d'une non-liaison et dans un cas seulement un autre pataquès a été produit, alors que les treize réparations de

pataquès des garçons ont autant été faites au profit de la bonne consonne de liaison (quatre), d'une non-liaison (quatre) que d'un autre pataquès (cinq ; Hornsby, 2012 : 77-78).

Ainsi, dans le test de lecture de Hornsby (2020), si les filles ont réalisé davantage de pataquès, malgré des données très limitées, elles pourraient avoir davantage conscience que leurs productions sont contraires à la norme, ce qui les amènerait à se corriger, quitte à opter pour l'option la plus sûre en ne réalisant pas les liaisons variables problématiques.

Nous venons d'aborder le phénomène de pataquès sous un angle d'approche sociolinguistique, en diachronie, en diamésie, en diatopie puis en diagénie, en rassemblant quelques observations de la littérature à ce sujet.

Dans cette partie, après avoir recensé les diverses appellations existantes pour mon phénomène d'étude et avoir opté pour le terme *pataquès*, j'ai passé en revue diverses définitions à lui donner avant d'opter pour celle de *liaison inattendue*. J'ai ensuite distingué les pataquès d'autres productions erratiques en matière de liaison, avant de présenter certaines caractéristiques linguistiques puis sociolinguistiques propres à ce phénomène.

Dans ce premier chapitre, nous sommes allés de la notion hyperonyme de sandhi à la notion très spécifique de pataquès, en nous arrêtant longuement en chemin sur la notion de liaison. Dans le chapitre suivant, je présenterai la méthodologie que j'ai mise en place dans le cadre de cette thèse.

Chapitre 2

Méthodologie

2.1 Méthodologie générale

2.1.1 État de l'art et collection d'exemples attestés

La première étape de la méthodologie que j'ai mise en place dans le cadre de cette thèse consistait à effectuer un va-et-vient entre état de l'art et collection d'exemples attestés.

J'ai effectué mes recherches de master (Coutanson, 2016, 2017) sur le phénomène de pataquès (comme indiqué en 3.2.1). Dès cette époque, j'ai entrevu que pour travailler sur un phénomène rare, il fallait tout d'abord recenser petit à petit toutes les appellations qui avaient pu être données à ce phénomène, et nous avions vu en 1.3.1 que celles-ci étaient très nombreuses pour le pataquès. En s'intéressant aux appellations communes à plusieurs auteurs, à l'aide des dates de publication et des références bibliographiques convoquées, on pourrait retracer comment un terme en particulier s'est répandu au sein de la communauté scientifique (Coutanson, 2020 : 2). Les fonctions *Prior works* et *Derivative works* du site Connected Papers (https ://www.connectedpapers.com/), un outil de visualisation de publications scientifiques similaires, permettent par exemple de retracer des lignages dans la littérature scientifique (voir figure 2.1).

Il est en outre nécessaire de bien définir le phénomène étudié, procédé complexe pour le pataquès comme nous l'avions vu en 1.3.2, et que pour cela il fallait avoir une bonne connaissance de l'état de l'art, en l'élargissant à des phénomènes connexes comme la liaison. La littérature portant sur le pataquès s'est révélée assez limitée. En effet, peu d'articles s'y consacrent intégralement (Léon, 1984 ; Desrochers, 1994 ; Hornsby, 2012) et il m'a alors fallu rechercher des mentions du phénomène dans des écrits portant sur la liaison ou la langue française en règle générale. De ce fait, une

recherche par mots-clés ou dans l'index d'ouvrages ne s'est pas toujours révélée très efficace. Des outils tels que *Connected Papers* auraient pu me faire gagner du temps si j'en avais eu connaissance plus tôt, bien que ceux-ci restent limités car ne référençant pas toute la littérature (voir à titre d'illustration les publications ayant un fort degré de similarité avec l'article de Desrochers (1994) : *Les Liaisons Dangereuses : Le Statut Équivoque des Erreurs de Liaison* dans la figure 2.1).

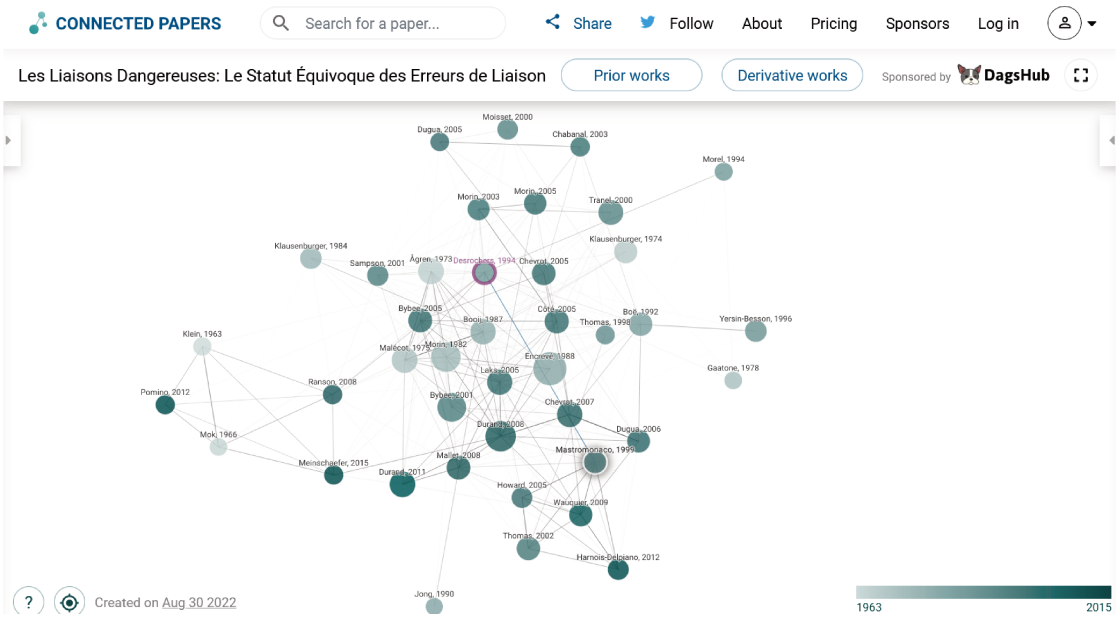


FIGURE 2.1 : Publications scientifiques similaires à l'article *Les Liaisons Dangereuses : Le Statut Équivoque des Erreurs de Liaison* (Desrochers, 1994)

Par ailleurs, le passage par une collection d'exemples semblait couler de source pour travailler sur un phénomène rare. De Fornel et Verdier (2018 : 5) définissent une **collection d'exemples** comme un « agrégat de cas singuliers » du phénomène que l'on souhaite étudier. Trouver des exemples oraux me semblait essentiel, et il fallait que ces exemples soient contextualisables, c'est-à-dire que je puisse connaître leur contexte de production afin d'éviter les occurrences inventées ou mal relevées. Dans la littérature, on peut douter de la véracité des exemples présentés hors contexte à titre d'illustration (voir notamment les exemples compilés par Desrochers, 1994 : 245-247, repris à plusieurs auteurs, sans réflexion aucune sur l'origine des données convoquées, autant issues de productions écrites (lettres adressées à l'Agence des Prisonniers de Guerre à l'époque de la Première Guerre mondiale) ou rapportées par des « puristes, littérateurs et amateurs » pour les données de Frei, 1929 : 12 que de productions orales (exemples

transmis de linguiste en linguiste et collections d'exemples non enregistrés, recueillis à la volée ou dans des chansons de tradition orale chez Morin & Kaye, 1982 : 323-325)). En ce qui concerne les exemples d'une collection, d'une part, ceux-ci peuvent être relevés à la volée ou être attestés. Si j'ai favorisé les exemples attestés qui peuvent être remobilisés avec plus de vigueur lors d'une argumentation, à l'image de tous les exemples présentés dans ce manuscrit, je ne me suis pas fermée aux exemples relevés à la volée, qui n'ont pas été enregistrés et qu'on ne peut certifier, qui nourrissent cependant ma connaissance du phénomène que j'analyse. D'autre part, une collection d'exemples peut être recueillie dans un même corpus ou dans des sources diverses. J'ai tout d'abord relevé des exemples de pataquès produits dans des situations d'énonciation variées, rencontrés dans mon temps libre ou dans le cadre de mes recherches, sans constituer une liste concrète mais plutôt en me familiarisant avec des exemples attestés, en conservant parfois le lien de l'audio où ils étaient produits. En parallèle, pour mes mémoires de master, j'ai amassé des exemples écrits issus de recueils de chansons et des exemples oraux issus d'une base de données ethnomusicologique (voir 3.2.1), ce qui m'a par exemple permis de connaître les contextes morphosyntaxiques préférentiels de réalisation de pataquès dans ces chansons ou encore de savoir quels pataquès étaient les plus et les moins prototypiques (Coutanson, 2020 : 4). Collecter des exemples dans un seul corpus permet de savoir si l'apparition de tel ou tel pataquès est liée à une certaine situation d'énonciation.

Pourtant, se limiter à une collection d'exemples présente un problème de taille : on ne peut faire que des statistiques sur les occurences attestées, mais aucune sur les occurrences non attestées, car on connaît les contextes dans lesquels elles ont été réalisées sans rien savoir des contextes dans lesquels elles n'ont pas été produites. De cette façon, si on peut observer à l'aide d'une collection d'exemples la part de réalisation de chaque consonne de pataquès dans l'ensemble des pataquès réalisés (pour un relevé systématique sur un même corpus), il est impossible d'obtenir un taux de réalisation de pataquès général en le rapportant au nombre total de contextes de liaison ou encore de frontières de mots. C'est là qu'une étude sur corpus se révèle intéressante, et je me suis lancée dans cette aventure dans le cadre de ma thèse. C'était également une opportunité au niveau théorique afin de regarder de plus près le lien qui existe entre pataquès et liaison.

Une question subsiste : une collection d'exemples peut-elle être dépassée pour l'étude de phénomènes aussi rares que le pataquès ? Une étude sur corpus est-elle envisageable dans le cadre d'une thèse, n'est-ce pas une perte de temps que de fastidieusement constituer un corpus pour ne relever que de maigres occurrences ? C'est la question que se

posent Cappeau et Gadet (2007b : 102) en mettant en avant l'exemple du pataquès :

> Faudra-t-il dès lors récuser toute autre méthode d'approche de données, comme la prise de notes à la volée ? Pourtant, y a-t-il d'autres moyens d'accéder à des phénomènes dont on ne peut pas savoir à l'avance si et quand ils vont émerger ? Ainsi d'actes de langage comme le rire ou les compliments[,] de variations morphologiques verbales, d'erreurs de liaison, etc,. que l'on rencontre en trop faible quantité dans les corpus oraux, qui deviennent finalement trop petits !

Pour étudier un phénomène rare, il faudrait en effet fouiller des données massives, mais l'état d'avancement du corpus détermine la faisabilité de cette entreprise : celle-ci sera facilitée si le corpus est déjà transcrit et encore davantage s'il est systématiquement annoté pour le phénomène étudié.

Historiquement enfin, les chercheurs souhaitant travailler sur des données authentiques, que ce soit les comparatistes européens, les structuralistes américains, les lexicographes, les dialectologues, les linguistes de terrain ou encore les acquisitionnistes, devaient souvent se limiter à des collections d'exemples, entendus ou lus. À l'écrit, avant l'avènement d'une linguistique de corpus informatisée, les corpus existaient, mais les analyses systématiques de ceux-ci demandaient des ressources humaines considérables et prenaient du temps, comme pour la constitution d'index et l'étude des concordances bibliques qui remontent au XIII^{ème} s. ou encore l'étude des concordances d'œuvres littéraires, notamment de Shakespeare (McCarthy & O'Keeffe, 2010 : 3). À l'oral, la captation des données est possible depuis un peu plus d'une centaine d'années seulement, mais cela ne veut pas dire que les premiers corpus étaient systématiquement transcrits et fouillés au-delà d'une collection d'exemples : c'est la fonction de réservoir d'exemples (Leech, 1991 : 13) à laquelle se limitent certains linguistes. McCarthy et O'Keeffe (2010 : 5) rappellent que "[i]t was the revolution in hardware and software in the 1980s and 1990s which really allowed corpus linguistics as we know it to emerge". L'amélioration de la puissance de calcul des ordinateurs, le développement d'internet et le téléchargement toujours plus rapide qui ont favorisé l'échange de données entre chercheurs, la miniaturisation des appareils et supports d'enregistrement audiovisuels (McCarthy & O'Keeffe, 2010 : 5) favorisant leur portabilité (Labov, 1976 : 283-284) ou encore l'océrisation de données textuelles massives ont largement contribué à la linguistique de corpus telle qu'on la connaît.

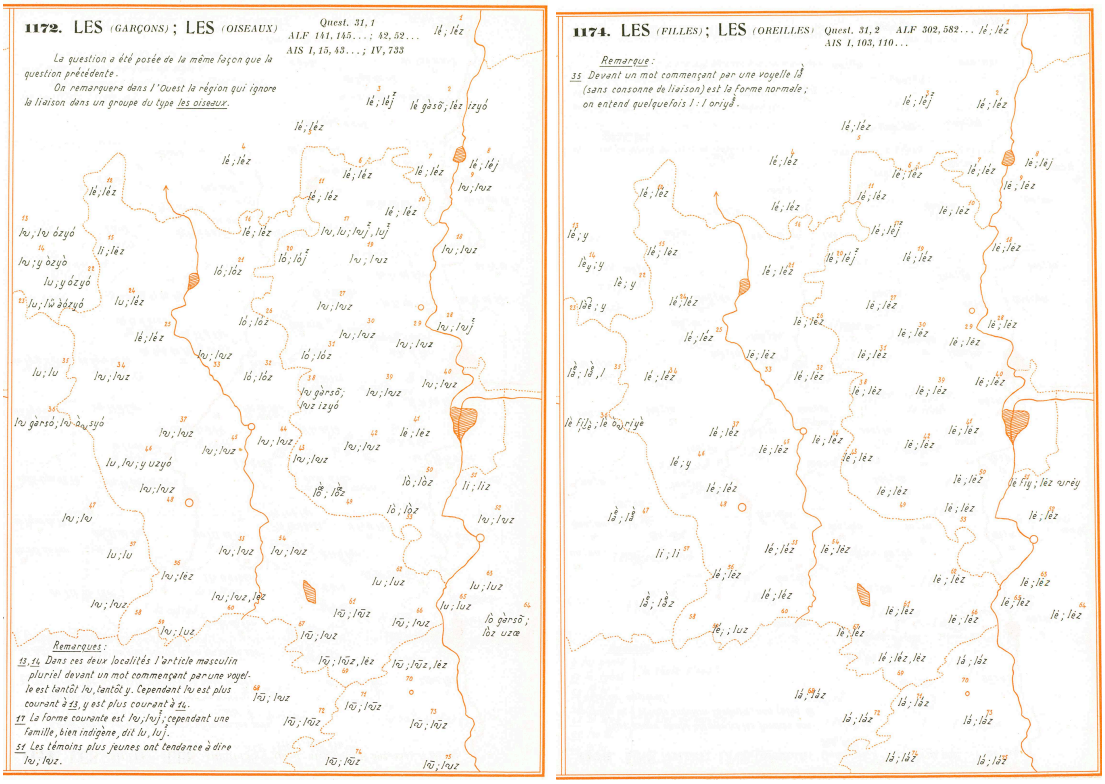
Pour résumer, je me suis rendue compte que cette compilation d'exemples en collec-

tion reprécisait petit à petit la définition adoptée de mon objet d'étude, elle-même influencée par ma revue de la littérature, cette dernière s'appuyant sur mon recensement des appellations données au phénomène et orientant les exemples collectés. De cette façon, le va-et-vient entre état de l'art et collection d'exemples attestés mis en place dans la première étape de cette méthodologie s'est révélé pertinent. Cette approche semble tout particulièrement adaptée à l'étude de phénomènes rares. D'une collection d'exemples débutée en master, je me suis orientée vers une étude sur corpus en thèse, afin d'éprouver la rareté du phénomène de pataquès et des différentes consonnes de pataquès et de comprendre le lien qui existe entre les phénomènes de pataquès et de liaison. Pour finir, j'ai souligné que les avancées technologiques des dernières décennies en matière de collecte et de traitement de données authentiques ont favorisé le passage de collections d'exemples à des corpus.

2.1.2 Choix et carottage de corpus disponibles variés

La seconde étape de cette méthodologie nécessitait une sélection de corpus disponibles variés, adaptés à l'étude du pataquès et accessoirement de la liaison, puis un carottage de ceux-ci.

Le choix d'un corpus découle tout d'abord de la ou des langues qui nous intéressent, dans laquelle ou lesquelles notre phénomène se produit. On gardera à l'esprit de jeter un oeil à ce qui se fait dans d'autres langues génétiquement ou typologiquement proches (Coutanson, 2020 : 4). En ce qui concerne la présente thèse, la langue à laquelle je m'intéresse est le français. J'ai dû pour commencer me demander si d'autres langues romanes présentaient des pataquès. Il fallait pour cela regarder de plus près lesquelles, à l'exception du français, présentaient des liaisons : celles-ci peuvent être observées dans les autres langues d'oïl (par exemple, en jersiais, un dialecte du normand ; cf Jones, 2012 ou en wallon de Liège ; cf Morin, 2005), en francoprovençal (Stich, 1998), voire dans certains dialectes occitans (observer par exemple les liaisons prénominales des points 58, 59 et 69 des cartes 2.2a et 2.2b, correspondant respectivement aux villages occitanophones d'Usson-en-Forez, d'Apinac et de Saint-Sauveur-en-Rue de l'*Atlas linguistique et ethnographique du Lyonnais*). Par ailleurs, on trouve des sandhis externes consonantiques non étymologiques en anglais (cf 1.1 et 1.2.1) ou encore en russe (Coutanson, 2020 : 4), deux autres langues indo-européennes. Ces cas ont également suscité mon intérêt, car d'un point de vue théorique la provenance de ces consonnes questionne. Enfin, il ne faut pas dédaigner les langues présentant plus généralement des sandhis externes consonantiques, car c'est un terrain favorable à l'émergence de cas inatten-



(a) 1172 : Les garçons, les oiseaux (b) 1174 : Les filles, les oreilles

FIGURE 2.2 : Cartes de l'*Atlas linguistique et ethnographique du Lyonnais*

dus, qui peuvent avoir été peu voire pas documentés. Pour les langues romanes, un passage par une étude sur corpus est envisageable. Par contre, pour les autres langues indo-européennes, il semble raisonnable de s'arrêter à l'étape de va-et-vient entre état de l'art et collection d'exemples attestés (Coutanson, 2020 : 4), pour tout simplement se familiariser avec ces phénomènes. Pour les langues dans lesquelles il n'y a que des sandhis attendus, un état de l'art est suffisant.

Il faut en outre opter soit pour un corpus synchronique, ce qui nécessitera de sélectionner une époque en particulier, soit pour un corpus diachronique/longitudinal, ce qui impliquera le choix d'une ou de périodes données. Les corpus longitudinaux sont utilisés en acquisition lorsqu'on se penche sur l'évolution du langage d'un enfant dans le temps ou du langage des apprenants lors de l'apprentissage d'une langue étrangère. Les corpus diachroniques et longitudinaux peuvent être des **corpus clos** ou **corpus fermés** (des ressources immuables, auxquelles ne sont pas ajoutés de documents supplémentaires) ou des **corpus de suivi** (corpus qui ne cessent de croître, « permettant l'analyse

chronologique, année par année par exemple, de données langagières » ; Habert et al., 1997 : 127, 147). Notons qu'en dehors d'un corpus diachronique, on peut tout autant parler de corpus clos, qu'on opposera plus généralement aux **corpus ouverts**, ou « corpus qui continuent à être enrichis » (Cappeau & Seijido, 2005 : 10), par exemple avec des données produites dans une situation d'énonciation différente. Dans mon cas, j'avais besoin d'un corpus synchronique pour l'étude du pataquès, afin d'examiner le fonctionnement de ce phénomène. J'ai par ailleurs constitué un corpus micro-diachronique avec le corpus des tops 1, mais l'objectif était plutôt ici d'étudier l'évolution de la liaison. Ce corpus est clos, mais il pourrait tout à fait devenir un corpus de suivi mis à jour chaque décennie.

Pour ce qui est de la modalité du corpus, on a le choix entre corpus écrit et corpus oral. Pour ce dernier type, les données peuvent être uniquement audio ou vidéo, ou encore multimodales (audio et vidéo ; Coutanson, 2020 : 5). On peut s'intéresser à la langue parlée, mais aussi à la langue chantée, à la voix criée, murmurée ou encore craquée et même à une langue signée, sifflée ou tambourinée. Ces différentes modalités détermineront s'il est nécessaire d'accéder à un corpus audio et/ou vidéo. Pour ma part, je recherchais des données orales et je m'intéressais au pataquès (et à la liaison) dans la langue chantée tout particulièrement, en conservant un intérêt pour la langue parlée également. De ce fait, un corpus audio m'était nécessaire au minimum.

Concernant le choix d'un **corpus de référence** (un « corpus qui propos[e] des variétés et des genres représentatifs d'une langue, incluant l'écrit et l'oral, des registres informels [et] soutenus, etc. » ; Tutin et al., 2019) ou d'un **corpus spécialisé** (un « corpus qui ne concern[e] que des usages spécialisés de la langue » ; Tutin et al., 2019), il faut se demander si on souhaite se concentrer sur un contexte énonciatif en particulier ou sur l'ensemble des contextes énonciatifs d'une langue. Pour cette thèse, j'ai déjà annoncé avoir voulu me limiter à des corpus oraux, mais l'état d'avancement des différents corpus existants ne m'a pas toujours convenu. En effet, bien que transcrits et annotés (par exemple, ils sont souvent lemmatisés et étiquetés morphosyntaxiquement), l'annotation de la liaison, que j'ai favorisée, reste propre à des corpus spécialisés, constitués pour l'étude de la liaison.

Quant au choix d'un **corpus brut**, c'est-à-dire présentant seulement des **données primaires** (des données linguistiques telles qu'elles ont été produites par le locuteur ou le scripteur) ou d'un **corpus annoté**, c'est-à-dire comprenant à la fois des données primaires et des **données secondaires** (des informations linguistiques ou extralinguistiques ajoutées par le linguiste au niveau d'éléments linguistiques, comme des descriptions, transcriptions ou annotations ; Baude et al., 2006 : 45), il existe un conflit entre

fidélité à la production linguistique originelle (avantage aux données primaires) et coût temporel et humain de traitement des données (avantage aux données secondaires) : si un corpus transcrit et annoté est préférable car son état d'avancement permet une analyse plus rapide, une mauvaise qualité ou l'absence de transcription et d'annotation peut amener à préférer un corpus brut ou seulement transcrit. Des données produites par d'autres annotateurs peuvent de plus présenter du bruit (dans mon cas, des occurrences considérées à tort comme pataquès ou comme liaisons) ou du silence (dans mon cas, des occurrences de pataquès ou de liaisons omises) : si le premier pourra être contrôlé, le second nécessitera un retour complet sur le corpus qui ne semble pas réalisable au regard de la masse de données nécessaire à l'étude d'un phénomène rare (Coutanson, 2020 : 5). La rareté d'un phénomène empêche de maintenir une attention constante lors de l'annotation, pouvant amener à des sous-catégorisations, et inversement, une focalisation trop importante peut conduire l'annotateur à des sur-catégorisations (Coutanson, 2020 : 5). Pour cette thèse, j'ai dû repartir des données primaires de certains corpus (des audios pour des données orales), j'ai pu effectuer une première fouille des données secondaires d'autres corpus (des transcriptions/annotations de données orales) notamment annotés pour la liaison, quand d'autres encore se sont révélés intéressants mais n'étant pas annotés pour la liaison (a minima) voire transcrits orthographiquement, le travail à fournir était trop important pour espérer leur fouille, tous ne pouvant pas l'être.

Abordons à présent la question de l'accès à des corpus linguistiques. La solution la plus simple est de réutiliser des corpus linguistiques librement accessibles en ligne. Ils peuvent être référencés au niveau national, comme pour la France sur la Plateforme d'Outils et de Ressources linguistiques pour un Traitement Optimisé de la LANGue française (ORTOLANG)[1], comportant des corpus écrits, oraux et multimodaux, ou encore sur la plateforme COllections de COrpus Oraux Numériques (COCOON)[2] pour des corpus uniquement audio (Coutanson, 2020 : 5). Au niveau européen, les corpus sont référencés sur la *Common Language Resources and Technology Infrastructure* (CLARIN)[3] (explorable grâce au *Virtual Language Observatory* (VLO)), tout comme sur l'*Open Language Archives Community* (OLAC)[4] au niveau international (Coutanson, 2020 : 5). Il faudra faire attention, car certains sites ne permettent qu'un accès partiel aux données ou qu'une interrogation du corpus (Cappeau & Gadet, 2007a : 130). Lors de cette étape

[1]Accessible à l'adresse suivante : https ://www.ortolang.fr/.

[2]Accessible à l'adresse suivante : https ://cocoon.huma-num.fr/.

[3]Accessible à l'adresse suivante : https ://www.clarin.eu/.

[4]Accessible à l'adresse suivante : http ://www.language-archives.org/.

de choix des corpus, j'ai par exemple sélectionné le corpus ALIPE, accessible sur ORTOLANG, ou encore l'ALAVAL, un atlas linguistique disponible sur un site spécifique.
Une autre façon d'accéder à des corpus est de réutiliser des corpus librement accessibles en ligne dans des banques de données non linguistiques. Il s'agit notamment de corpus issus d'autres disciplines ou d'archives patrimoniales (Coutanson, 2020 : 6). Ces autres disciplines seront par exemple : « l'ethnologie, l'anthropologie, la sociologie, la psychologie, la démographie, l'histoire orale [...] [qui] utilisent l'enquête orale, le témoignage, l'interview, le récit de vie » (Baude et al., 2006 : 19). Pour ce qui est des archives patrimoniales, on peut par exemple citer au niveau national les archives sonores ou les vidéos de la Bibliothèque nationale de France (BnF), accessibles sur le portail Gallica[5] ou encore les archives de l'INA[6] (Coutanson, 2020 : 6). En ce qui me concerne, j'ai par exemple sélectionné le FJD et le fonds Charles Joisten sur la Base inter-régionale du patrimoine oral (BIPO), une base de données patrimoniale, notamment ethnomusicologique.
Seront encore plus difficiles d'accès des corpus scientifiques non partagés avec les pairs ou tout au moins restreints à une utilisation en interne au sein d'un groupe de recherche. En 2007, selon Cappeau et Gadet (2007a : 132), les corpus oraux les plus nombreux pour le français auraient notamment été recueillis dans le cadre de mémoires ou de thèses, et auraient moins été diffusés. On peut espérer qu'aujourd'hui les corpus de masterants et doctorants soient de plus en plus rendus disponibles après la recherche, mais pour les corpus des générations précédentes ainsi que pour les corpus encore non publiés, il est toujours possible d'effectuer une demande afin d'y accéder, mais celle-ci peut ne pas aboutir. L'absence de métadonnées ou d'une description du corpus peut aussi limiter l'analyse, voire être rédhibitoire pour le chercheur. Ce cas de figure s'est présenté à moi pour le corpus PFC, qui bien que partiellement accessible sur ORTOLANG, n'est disponible en totalité que sur un site en accès restreint pour les chercheurs en possédant l'accès, faute d'anonymisation de certaines transcriptions. J'avais demandé un accès dès le master, mais je n'ai eu connaissance de l'existence de mon compte qu'en début de doctorat.
Quand le corpus que l'on souhaite obtenir n'existe pas encore, il reste possible de compiler des documents en corpus, la tâche pouvant être complexifiée si ceux-ci sont des documents privés. Dans le cadre de cette thèse, j'ai compilé moi-même le corpus dit de tops 1, regroupant les chansons ayant accédé à la première place de classements musicaux pour la France entre 1956 et 2017.

[5]Accessible à l'adresse suivante : https ://gallica.bnf.fr/.
[6]Accessible à l'adresse suivante : https ://www.ina.fr/.

L'ultime solution, plus complexe encore, est de recueillir soi-même un corpus lorsque les données telles qu'on les recherche font défaut, ce qui sera abordé dans la partie suivante (2.1.3). Je n'ai pas eu à recueillir un nouveau corpus oral pour cette thèse. Parfois, le recueil d'un corpus peut être imposé par un cadre méthodologique, afin de faire corps avec son objet d'étude. Dans ce cas-là, l'étape de carottage que je vais à présent aborder peut se révéler superflue.

Dans le cadre de ce travail de recherche portant sur un phénomène rare, nous avons, avec mon directeur de thèse, opté pour un carottage préalable à l'étude sur corpus. Il s'agissait d'un essai, que je commenterai dans la partie analyse (4.1). Arrêtons-nous donc sur cette notion de **carottage**, notion encore peu répandue dans le champ de la linguistique et en cours de définition et de délimitation. Baude et Dugua (2015 : 358) indiquent que :

> [l]e terme carottage est utilisé dans un sens métaphorique pour définir une méthodologie empruntée à d'autres disciplines et qui consiste à analyser des échantillons du corpus à travers les différentes strates qui le constituent. Nous considérons en effet chacune de ces micro-analyses fondée sur des échantillons (carottes), comme une possibilité de sonder un grand corpus non pas à différents « endroits » mais plus exactement sous différents angles et par là, différentes perspectives théoriques.

On pourrait se demander en quoi un carottage diffère d'un échantillonnage. En linguistique de corpus, un échantillonnage peut être mis en place à deux moments : lors de la constitution d'un corpus équilibré ou lors d'une première analyse d'un gros corpus, afin de réduire la masse de données étudiées. Dans le premier cas, Habert et al. (1997 : 149) rappellent qu'il faut veiller à équilibrer le corpus en fonction du type de données langagières à intégrer (par exemple, selon leur modalité, leur genre, la période à laquelle elles ont été produites ou encore les caractéristiques sociologiques des locuteurs pour des données orales ; Desagulier, 2017 : 3 ; Habert et al., 1997 : 150) et de leur taille, sans sur-représenter certaines parties des documents retenus (par exemple l'introduction, dans un corpus écrit constitué d'articles). Pour le second cas, lors d'un échantillonnage d'un corpus a posteriori, il faut soit sélectionner aléatoirement certains textes, énoncés… à analyser d'un corpus équilibré, soit effectuer une sélection équilibrée dans une base de données (selon les mêmes critères). Desagulier (2017 : 3) fait ce rapprochement très parlant à propos de l'échantillonnage :

> [l]ike astrophysics, which infers knowledge about the properties of the universe and beyond from the study of an infinitesimal portion of it, corpus

> linguists use finite portions of language use in the hope that they will reveal the laws of a given language, or some aspect of it.

Baude (2018, communication personnelle) estime que : « la [différence majeure] est qu'un échantillonnage répond à un objectif de représentativité [...] tandis qu'un carottage [...] a pour objectif de donner des informations sur un « objet »[,] sans implication sur la représentativité de cet objet ». De cette façon, un carottage permet de se faire une première idée d'un corpus lorsqu'on ne sait pas ce qu'on peut y trouver, tout en sachant pertinemment que ce que l'on observe n'est pas représentatif du corpus étudié. Le carottage est utilisé comme une étape préparatoire à l'étude sur corpus, pour savoir s'il faut poursuivre dans ce sens. J'ai utilisé cette méthode afin de déterminer quels corpus étaient les plus prometteurs.

Au cours de cette étape, je suis tout d'abord revenue sur le processus de sélection d'un corpus linguistique, en distinguant les corpus monolingues des corpus multilingues, les corpus synchroniques des corpus diachroniques ou longitudinaux, les corpus clos des corpus de suivi ou ouverts, les corpus écrits des corpus oraux, les corpus de référence des corpus spécialisés et enfin les corpus bruts des corpus annotés. Pour ma part, j'étais en quête de corpus oraux spécialisés (pour l'étude de la liaison), en français (voire dans une autre langue d'oïl, en francoprovençal ou en occitan), concernant la langue chantée (voire parlée). J'ai opté pour des corpus bruts, l'un synchronique pour l'étude du pataquès et l'autre micro-diachronique afin de me pencher sur le phénomène de liaison que j'ai par la suite annotés, tout en effectuant une première fouille de corpus annotés pour la liaison. Dans cette partie, nous avons également vu comment le linguiste peut accéder à des corpus, que ce soit en réutilisant des corpus linguistiques (accessibles dans des bases de données nationales, européennes ou internationales), issus d'autres disciplines, d'archives patrimoniales ou encore des corpus scientifiques restreints à un groupe de recherche ou non partagés, en compilant des données accessibles ou privées en corpus, voire en recueillant soi-même un corpus. J'ai décidé de sélectionner cinq corpus durant cette étape : l'ALAVAL, un atlas linguistique, le corpus ALIPE, le fonds Charles Joisten, le FJD et le corpus PFC. J'ai terminé en faisant appel à la notion de carottage en linguistique, qui correspond à un échantillonnage non représentatif d'un corpus, donnant des informations au chercheur sur cet objet.

2.1.3 Constitution de nouveaux corpus

Suite au va-et-vient entre état de l'art et collection d'exemples attestés, après avoir choisi puis carotté divers corpus, il est possible de passer par une étape de constitution de nouveaux corpus, soit en compilant des données déjà existantes, soit en recueillant un nouveau corpus, comme mentionné en 2.1.2. Quels sont les corpus qui font défaut ? Selon Habert, en 2000, les variétés de français sous-représentées en linguistique de corpus étaient les suivantes : le français non hexagonal, le français scientifique et technique et le français non standard, alors que les variétés sur-représentées étaient les suivantes : les chronolectes antérieurs au français contemporain, le français écrit et le français suivant les règles du bon usage (Habert, 2000 : 4). Nous allons passer en revue de nombreuses données que le linguiste peut compiler ou recueillir en corpus. Ces données peuvent être plus ou moins spontanées ou **élicitées**, c'est-à-dire provoquées par le chercheur. Par ailleurs, selon les cas, elles relèvent du domaine public ou privé. Le moins impliquant pour un chercheur est de compiler des données déjà existantes en corpus (Coutanson, 2020 : 16).

Si on a besoin d'un nouveau corpus écrit, on le constituera souvent par compilation, les documents, produits avec une autre finalité, préexistant à la recherche linguistique. Par exemple, on composera un corpus de copies d'étudiants ou d'évaluations scolaires, de dossiers rédigés par des étudiants, de mémoires ou de thèses, de notes de cours, d'articles de presse ou scientifiques, de livres, de comptes-rendus de réunion, de lettres, d'SMS, de mails, de publications sur les réseaux sociaux, de commentaires de publications internet, de messages instantanés sur des chats, de messages de forums, de pages web, de textes de lois, de transcriptions de discours ou encore de modes d'emploi (Coutanson, 2020 : 11). Il est tout de même possible de recueillir certains corpus écrits, comme en demandant aux étudiants une production écrite répondant à une consigne spécifique. Il sera alors question de données élicitées.

Si on a besoin de constituer un corpus oral à présent, notons tout d'abord qu'on peut souhaiter obtenir des données monologales ou des données dialogales. On peut à nouveau compiler des données déjà existantes, produites avec une autre finalité que l'étude linguistique, qu'il s'agisse : de livres audio, de chansons, de données télévisuelles ou radiophoniques, de podcasts, de MOOC, de vidéos en ligne, d'annonces sonores diffusées dans les lieux publics, de messages vocaux laissés sur un répondeur téléphonique, de vocaux partagés sur des messageries instantanées ou encore d'enregistrements sonores familiaux (Coutanson, 2020 : 11). Pour l'oral cependant, il faudra bien souvent enregistrer des productions qui n'ont pas vocation à l'être dès l'origine, comme : des cours, des

conférences, des récitations (poésies, comptines, théâtre, film…), des contes ou chants de tradition orale, des réunions professionnelles, des sermons, des interactions (médiées ou non par téléphone/ordinateur, privées ou professionnelles), des entretiens, des récits de vie ou de faits divers, une tâche de lecture, des descriptions suite à des stimuli visuels, sonores ou olfactifs ou encore une tâche de traduction (Coutanson, 2020 : 11 ; Baude et al., 2006 : 27), afin d'obtenir un corpus faisant encore défaut. Les trois derniers types de données tout particulièrement, mais pas seulement, sont élicités.

Par conséquent, lors de la constitution d'un nouveau corpus, des données écrites ou orales, publiques ou privées peuvent être compilées ou recueillies. Celles-ci sont à replacer sur un continuum entre spontanéité et élicitation. Dans cette sous-partie, j'ai mentionné de multiples exemples de données linguistiques. Les données écrites sont plus souvent compilées que recueillies, ce qui accélère grandement cette phase de constitution de corpus. Au contraire, les données orales, qui peuvent être monologales ou dialogales, sont couramment captées par le chercheur lui-même lorsqu'un nouveau corpus se révèle nécessaire, ce qui est à la fois impliquant pour le chercheur et chronophage. Je vais désormais m'arrêter longuement sur la chaîne de traitement de mes deux corpus de chansons. Afin de ne pas me répéter, j'ai regroupé ici la chaîne de traitement du FJD, déjà constitué en fonds et provenant d'une autre discipline, l'ethnomusicologie, et celle d'un corpus que j'ai compilé au cours de cette étape, le corpus de tops 1.

2.1.3.1 Obtention d'un fichier son et mise en place d'un plan de nommage

Lors de cette première étape, il faut obtenir un enregistrement audio et établir un plan de nommage.

Comme vu en 2.1.2, on peut se procurer le document sonore en enregistrant soi-même un corpus, en compilant des données déjà existantes afin de constituer un corpus ou en réutilisant les données primaires d'un corpus établi. Dans le premier cas, il faudra alors créer des données primaires et des données secondaires alors que dans les derniers cas il ne sera question que de données secondaires : c'est ce que j'ai eu à produire avec mes deux corpus de chansons. J'y reviendrai dans la méthodologie de chaque corpus.

En ce qui concerne le plan de nommage, dans les deux premiers cas, il faudra soi-même en mettre un en place tandis que dans le dernier cas, les appellations des données primaires pourront être reprises lors de la création de données secondaires.

Il sera donc nécessaire d'obtenir un enregistrement audio en élaborant soi-même un corpus, en compilant des données existantes en corpus ou en réutilisant un corpus et d'instaurer un plan de nommage ou de réemployer les appellations des données primaires pour les données secondaires.

2.1.3.2 Préparation d'un fichier .cha

La deuxième étape est de créer un fichier .cha, c'est-à-dire de préparer un fichier pour travailler sur le logiciel *Computerized Language ANalysis* (CLAN) (MacWhinney, 2000a), en utilisant le même identifiant que l'audio qui y sera associé. CLAN est un logiciel très utilisé par les chercheurs en linguistique dans les domaines de l'acquisition du langage, de l'analyse conversationnelle, du bilinguisme, de l'apprentissage d'une langue seconde ou encore des atypies langagières (MacWhinney, 2000a : 5). Il faut bien distinguer CLAN, une « série de programmes informatiques qui permettent l'alignement [de] transcriptions avec [des audios ou des] vidéos et de traiter et analyser [des] données automatiquement », de CHAT (.cha), un « format de transcription et de codage qui permet d'uniformiser [les corpus produits avec CLAN] » (standardisés donc) et de la banque de données associée qui rend accessibles des corpus réalisés à l'aide de CLAN (Bourdoux et al., 2011 : 3). Ces trois éléments sont compris dans CHILDES, un « système d'échange et de description du langage de l'enfant » (Bourdoux et al., 2011 : 3), devenu aujourd'hui le projet Talkbank (MacWhinney, 2000a : 12).

Pour sa thèse de doctorat, Loïc Liégeois (2014) a utilisé CLAN afin de transcrire et d'annoter le premier corpus acquisitionnel pour l'étude de la liaison en français. Par la suite, une équipe de chercheurs s'est réunie pour réfléchir à une annotation commune des liaisons dans l'environnement langagier des enfants (Dugua et al., 2017). Travaillant majoritairement dans le domaine de l'acquisition du langage, ces chercheurs ont logiquement opté pour l'utilisation de CLAN, contrairement aux pratiques d'autres équipes de chercheurs étudiant la liaison comme les membres du projet PFC qui se servent du logiciel Praat[7] (Boersma & Weenink, 1992). Le 24 janvier 2019, nous avons refait une réunion avec certains membres du groupe de recherche (Olivier Baude, Céline Dugua, Jennifer Ganaye, Flora Badin, Christophe Parisse et Loïc Liégeois). Nous avons poursuivi la réflexion sur l'annotation des liaisons en corpus, en conservant CLAN pour la transcription et l'annotation de nos données. J'avais reçu en master une rapide formation à ce logiciel pour travailler sur des données acquisitionnelles. En débutant ma thèse, je pensais que cette formation était suffisante et je ne m'étais pas inscrite à la

[7] du néerlandais 'parle'

formation à CLAN organisée par le consortium Corpus, langues, interactions (CORLI), à laquelle je n'ai participé qu'en 2019. Mais c'est surtout en me replongeant dans les manuels de CLAN (MacWhinney, 2000a,b), en rédigeant un tutoriel pour mon usage personnel et en m'entraînant sur mes propres données que ma maîtrise du logiciel s'est améliorée.

Chaque fichier .cha est borné au début et à la fin par @Begin et @End. Le document comprend en premier lieu une[8] en-tête qui va apporter des informations à son propos : il s'agit des métadonnées. Pour moi, il s'agit d'apporter des informations sur chaque chanson, comme quelques informations d'ordre sociolinguistique sur le chanteur. J'indiquais les informations suivantes : l'âge (indiqué au format année ;[9]) et le sexe présumé du chanteur (Male pour un individu de sexe masculin ou Female pour un individu de sexe féminin[10]) ainsi que son rôle : Singer. S'il y avait plusieurs interactants, les informations de chacun étaient ajoutées. Le rôle Singer n'est pas un rôle par défaut sous CLAN. Il ne faut donc pas oublier de le déclarer dans le depfile. Selon le corpus, d'autres informations pertinentes étaient ajoutées. Celles-ci seront présentées pour chaque corpus. Dans les en-têtes sont aussi indiqués la langue du document sonore transcrit (uniquement le français car c'est la seule langue sur laquelle j'ai travaillé en corpus[11]) et le nom de la transcriptrice (j'ai été la seule à transcrire ces données). Enfin, @Media permet d'aligner le son avec le texte. Il suffit d'indiquer le nom du fichier son, qui doit se situer dans le même dossier que le fichier .cha, et de préciser *audio*, avant de sauvegarder le tout (seuls des audios sont associés aux fichiers .cha pour mes corpus de chansons).

Dans cette deuxième étape, j'ai tout d'abord effectué une brève présentation du logiciel CLAN, avant de le distinguer du format de transcription et d'annotation CHAT et de la banque de donnée qui leur est associée, tous regroupés dans CHILDES, aujourd'hui nommé le projet TalkBank. Ensuite, j'ai expliqué que j'utilisais ce logiciel suite

[8] Choix délibéré

[9] CLAN est un logiciel qui a longtemps été réservé à l'acquisition du langage. Normalement, l'âge du locuteur est indiqué selon le format année ;mois.jours, mais ici, avec une population principalement adulte, seul le nombre d'années est pertinent et donc conservé.

[10] CLAN propose également la catégorie Unknown. Pour le FJD, ma catégorisation se fonde sur les prénoms des chanteurs et il semble difficile d'obtenir plus d'informations à ce sujet. Pour les Tops 1, de la même façon, les informations nécessaires ne sont pas disponibles ou sont difficilement accessibles (par exemple, sur la base d'auto-déclarations en interview), ce qui m'oblige à me cantonner à cette dualité.

[11] Les langues sont indiquées sous CLAN sous la forme de codes au format ISO 639, proposés par l'Organisation internationale de normalisation.

à la volonté d'une équipe de chercheurs travaillant sur des corpus d'étude de la liaison d'uniformiser l'annotation de ce phénomène, et que pour le traitement des corpus utilisés dans le cadre de ma thèse, j'ai dû approfondir mon maniement de ce logiciel. Pour terminer, j'ai mentionné des éléments communs aux métadonnées à renseigner pour les fichiers .cha créés : âge, sexe présumé et rôle de l'interactant ; langue, nom du transcripteur et enfin fichier son à associer afin de préparer l'alignement son/texte, étape que nous allons aborder dans le point méthodologique suivant.

2.1.3.3 Alignement son/texte

Pour aligner le son avec le texte dans cette troisième étape, un fichier son doit avoir été associé au fichier .cha au préalable, comme présenté en 2.1.3.2. Pour travailler sur un document à découper à intervalles réguliers comme une chanson, je suis partie de mon fichier avec les métadonnées renseignées, j'ai activé la lecture à l'aide de la touche F5 puis j'ai balisé mon document sonore à l'aide de la touche espace, avant d'enregistrer les modifications. En cas d'erreur, il suffit de repartir de la dernière balise correcte et de reprendre la découpe.

Le corps d'une transcription effectuée sous CLAN est fragmenté en tires. Une **tire** correspond à une portion de transcription associée à une balise initiale et à une balise finale. Étant donné qu'une chanson prend d'ordinaire une apparence monologale, on pourrait être tenté de mettre dans une seule tire une portion de chant sans interruption, voire la chanson complète. Cependant, pour les chansons il peut être plus avantageux d'opter pour une subdivision, sinon lors de la recherche d'une occurrence toute la chanson apparaîtra dans les résultats, ce qui alourdit leur lecture et ne facilite pas l'extraction des informations pertinentes.

À quelles unités linguistiques doivent coïncider les tires ? Généralement, sous CLAN, la production linguistique est découpée en tours de parole. Je n'ai pas choisi le même critère pour les deux corpus de chansons. Je présenterai donc ces critères dans la partie méthodologique propre à chaque corpus.

Pour cette étape, il est nécessaire de lancer l'audio depuis le fichier .cha avec les métadonnées renseignées, avant de baliser le document. Les chansons doivent bien être divisées en unités plus petites même s'il s'agit d'une intervention unique d'un seul participant. Une fois cette étape terminée, le gros du travail débute avec l'étape de transcription/annotation.

2.1.3.4 Transcription et annotation du document sonore

La quatrième étape consiste à transcrire et à annoter l'audio. Nous allons tout d'abord nous pencher sur la transcription et l'annotation sous CLAN, avant de nous intéresser plus particulièrement à l'annotation de la liaison et du pataquès, en présentant un outil qui a été élaboré afin de faciliter cette tâche.

2.1.3.4.1 Transcription et annotation sous CLAN

Dans un premier temps, la transcription et l'annotation se font sur le document .cha préparé, tire par tire, selon les règles de transcription et d'annotation élémentaires de CLAN (présentées dans MacWhinney, 2000a), soit par exemple :

- tires principales (destinées à la transcription de l'audio) précédées d'une[12] astérisque (‹*›), tires dépendantes (destinées à apporter des informations complémentaires sur la transcription) précédées du symbole pour cent (‹%›)
- absence de majuscules sauf pour les noms propres
- nombres écrits en lettres
- pas de ponctuation, excepté un marqueur de fin d'énoncé (à placer en fin de tire), qui est obligatoire. Par défaut, celui-ci est ‹.›. ‹?› est utilisé pour les questions et j'ai réservé ‹!› aux impératifs. Les apostrophes sont maintenues. Pour les clitiques verbaux, ceux-ci sont séparés de leurs hôtes par un espace au lieu d'un trait d'union. Les traits d'union des composés et des autres clitiques sont remplacés par des ‹+› : *là+bas*, *Marly+gomont*... Seule une majuscule initiale est utilisée pour les composés qui sont des noms propres. Cette dernière notation (à l'aide d'un ‹+›) pose problème lors des étapes de lemmatisation et d'étiquetage morphosyntaxique, car les mots graphiques ne sont pas reconnus. Comme tout ‹+› ne correspond pas à un trait d'union (‹-›), on ne peut pas les remplacer automatiquement par des traits d'union.
- transcription en orthographe standard[13], avec indication au besoin d'une prononciation non standard en API en commentaire :

(67) *SGR : eh@i Bonapart(e) apaise tes canons nous te ferons contribution.
%pho : e bonapaʁt apɛzə te kanɔ̃ nu tə fəʁɔ̃ kɔ̃stʁibysjɔ̃
(JD_a330_22, 43800, Henriette Usson)

[12] Choix délibéré

[13] Cela facilite à la fois la lecture (Baude et al., 2006 : 30), tout en permettant une bonne reconnaissance des mots par l'outil informatique.

- utilisation d'‹xxx› pour ce qu'on ne parvient pas à transcrire, d'‹yyy› lorsqu'on entend ce qui est prononcé mais qu'on ne le comprend pas (la prononciation est alors indiquée sur une ligne de transcription phonétique dépendante) :

(68) *SGR : brûleront yyy pilleront vos maisons.
%pho : brylərɔ̃ m m m pijərɔ̃ vo mɛzɔ̃
(JD_a330_22, 43800, Henriette Usson)

et de ‹www› lorsqu'on décide de ne pas transcrire un passage. J'ai utilisé cette dernière notation pour le code-switching (alternance codique), le passage à une autre langue que je ne transcrivais pas orthographiquement afin d'éviter un mauvais repérage de potentiels contextes de liaison par l'outil informatique ; j'indiquais alors la langue sur une tire dépendante explicative %exp : :

(69) *SGR : donne moi !
*SGR : www.
%exp : deux vers en latin.
(Tops 1, Enigma, *Sadeness part I*, 1991 [1990])

(70) *SGR : et www de ta voix câline.
%exp : vers partiellement en italien.
(Tops 1, Dalida, *Bambino*, 1957 [1956])

Dans le FJD, les chanteurs alternent parfois entre français et langue régionale (il semblerait qu'il soit question d'occitan le plus souvent).

- événements para-linguistiques indiqués en anglais au *simple present* à la troisième personne du singulier, précédés de ‹& = ›, comme ‹& = laughs› lorsque le locuteur rit ou ‹& = coughs› lorsqu'il tousse
- phonèmes non réalisés indiqués entre parenthèses (cas d'élisions) : ‹()›. Cela concerne principalement les schwas non réalisés en français (cf exemple 67), mais également d'autres voyelles, des consonnes ou plusieurs segments adjacents non réalisés. Les mots tronqués sont eux aussi notés de la sorte : *kalasch(nikov)*, *accro(chée)*. En ce qui concerne les mots graphiques complètement omis, ceux-ci ne sont tout simplement pas transcrits[14], comme dans les exemples 71 et 72 où les noms *esclave* et *blanche* ne présentent pas de déterminants.

[14] Le manuel de CLAN (MacWhinney, 2000a) indique que ces mots doivent être annotés avec un ‹0› les précédant, comme par exemple *0un esclave* et *0la blanche* pour les exemples 71 et 72. Ayant commencé à les annoter de la sorte, j'ai par la suite abandonné cette pratique suite à une discussion avec Christophe Parisse. En effet, cela alourdit la lecture des transcriptions et cette absence peut être détectée par exemple lors d'une extraction de concordances.

(71) *SGR : esclave n'a pas d(e) remise de peine.
(Tops 1, Booba, *DKR*, 2016 [2016])

(72) *SGR : mon département aime blanche de Bolivie.
(Tops 1, Booba, *É.L.É.P.H.A.N.T*, 2016 [2016])

J'ai conservé les apostrophes uniquement en cas d'élision systématique : ex : *l'amie*, *j'embrasse*[15]... Pour les autres voyelles ou devant initiale consonantique, j'ai transcrit les deux mots, dont un avec une élision : *d(e) soleil*, *s(i) il*... car les élisions ici ne sont pas systématiques.

- distinction des interjections, notées avec ‹x@i› :

(73) *SGR : oh@i promets moi d'être fidèle je te sortirai du couvent.
(JD_a075_23, 43810, Marie Soulier)

des interjections propres au chant, notées avec ‹x@si› :

(74) *SGR : médecin va t'en au diable oh@si !
*SGR : ah@si ah@si ah@si ah@si ah@si ah@si ah@si ah@si ah@si.
*SGR : moi que j'aime tant le vin.[16]
(JD_a330_04, 43800, Henriette Usson)

Les onomatopées sont pour leur part notées à l'aide de ‹x@o›. Lorsque qu'on veut bien expliciter qu'on se réfère à une lettre de l'alphabet mentionnée seule, on utilise ‹x@l› :

(75) *SGR : sans toi mon corps sous [z] x[17]@l.
(Tops 1, Mylène Farmer, *Oui mais... non*, 2010 [2010])

[15] Lors de mes premiers contacts avec CLAN, il nous avait été conseillé de transcrire entièrement ces mots : *j(e) embrasse*, *l(a) amie*... afin que le premier mot ne soit pas ambigu pour CLAN (par exemple *l'* peut correspondre à *la* ou *le*) au cours d'un calcul de fréquence des termes employés par exemple. J'ai à nouveau abandonné cette pratique suite à une discussion avec Christophe Parisse. Une fois de plus, cela alourdit la lecture des transcriptions et ce ne semble pas nécessaire au regard de la qualité des outils de lemmatisation actuels.

[16] J'ai au départ opté pour une notation des interjections propres au chant sous forme de blocs du type *ah* + *ah* + *ah@si*. J'ai ensuite décidé de casser ces blocs en éléments redondants (en *ah@si ah@si ah@si* par exemple), en pensant qu'une future analyse de ces interjections en serait facilitée. La première notation permettait de retrouver des séquences similaires dans diverses chansons, mais empêchait une analyse systématique à un niveau inférieur. Mais si les blocs ne se laissaient pas diviser en plusieurs interjections similaires, je ne les ai pas forcément séparés. Un chercheur qui travaillerait sur ce sujet serait amené à réviser le repérage et l'annotation de ces phénomènes pour une fouille la plus optimale possible des données.

[17] Cet exemple est piégeux car ici il est bien question de la lettre ‹x›.

et si plusieurs lettres sont épelées sans que cela corresponde à un sigle connu, on utilise ‹x@k›
- notation des sigles conventionnels en lettres majuscules avec un tiret bas entre chaque lettre : *A_K* (arme)...

J'ai fait le choix de ne pas annoter les pauses et les chevauchements, afin de ne pas alourdir la transcription.

Il reste maintenant à associer chaque tire à un participant. Cela peut être fait avant, pendant ou après la transcription. Sous CLAN, chaque tire est précédée d'un code composé de trois lettres et/ou chiffres pour chaque participant. Par exemple, pour les chansons, c'est prototypiquement le chanteur qui s'exprime et on lui associe le code SGR (voir figure 2.5 p.162). Pour associer chaque tire à un participant, il suffit d'utiliser le raccourci clavier associé à chacun (disponibles dans Tiers). En cas de difficulté, il faut vérifier que les participants qui ont été entrés lors de la deuxième étape (voir 2.3.2) aient bien été mis à jour au préalable à l'aide de Tiers\Update. J'avais préalablement pris la décision (voir 2.1.3.3) de découper les chansons en unités plus petites. Il faut alors bien rappeler le code associé au chanteur à chaque tire pour s'assurer de l'identité de la personne qui s'exprime lors d'une extraction de données pendant la fouille du corpus.

Au cours de cette étape, l'audio est transcrit et annoté en respectant les conventions générales du logiciel CLAN. J'ai mentionné certains retours sur les annotations auxquels j'ai procédé (concernant les mots omis, les élisions obligatoires de voyelles devant voyelles en français et les interjections propres au chant). Enfin, chaque tire doit être associée au code d'un participant de façon semi-automatique. Arrêtons-nous à présent sur l'annotation de la liaison et du pataquès mise en place pour ces deux corpus.

2.1.3.4.2 Annotation de la liaison et du pataquès

Lors des rencontres de l'équipe de chercheurs travaillant sur l'acquisition de la liaison, un système de codage des liaisons a été convenu (Dugua et al., 2017), comme mentionné en 2.1.3.2. Ce système de codage, que j'ai adopté, est le suivant (Dugua & Ganaye, 2022) :

- l'annotation se fait directement sur la ligne de transcription (les lignes principales de CLAN)
- l'annotation se fait entre crochets. Il s'agit ici d'une annotation propre à CLAN

et non d'une indication phonétique : "[...] CLAN considers that the material in square brackets refers to a single preceding word [...]" (MacWhinney, 2000a : 69)

- la consonne entre crochets renvoie à la consonne de liaison produite : [z], [n], [t], [r], [p] ; [o] est utilisé pour les liaisons non réalisées (cf Liégeois, 2014 : 185).

Pour les besoins de mon étude, j'ai dû adapter en partie ce système de codage :

- [x] se réfère toujours à une liaison réalisée avec x = consonne de liaison
- [ox] se réfère à une liaison non réalisée avec x = consonne de liaison qui aurait pu être réalisée (j'indique pour ma part la nature de cette consonne de liaison potentielle)
- [*x] se réfère à un sandhi externe inattendu (en particulier à un pataquès) avec x = consonne de sandhi ou de pataquès
- [x*] correspond à un cas incertain (perceptiblement, il est impossible de trancher entre liaison réalisée ou liaison non réalisée) ou à une consonne finale réalisée, habituellement non prononcée (comme dans les exemples 96 p.167, 113 p.171 ou encore 115 p.171), avec x = consonne de liaison ou consonne finale[18].

Le système de codage adopté pour cette thèse peut être mobilisé dès l'étape de transcription/annotation, toutefois ce n'est pas nécessaire car comme nous allons bientôt le voir, ces liaisons peuvent être préannotées automatiquement à partir de la transcription puis vérifiées manuellement afin de gagner du temps. La seule annotation obligatoire des liaisons dès l'étape de transcription concerne les sandhis externes inattendus hors contexte de liaison, qui ne pourront pas être retrouvés par l'outil informatique mentionné ci-dessous (comme pour le pataquès de l'exemple 45 p.102). Il est très important de noter ici que dans la chronologie de la méthodologie mise en place, il est indispensable de passer à ce stade à la dernière étape de réécoute et de vérification du fichier .cha produit (2.1.3.5). Les paragraphes suivants concernent également la phase d'annotation de la liaison et des pataquès dans les corpus et ont ainsi été regroupés avec cette étape. Pourtant, il est essentiel de fournir à notre outil des fichiers .cha finalisés.

Notre système de codage de départ se voulait, contrairement aux systèmes de codage utilisés pour l'annotation de la liaison dans les corpus ALIPE, HPOL1 et PFC, minimal (Dugua & Ganaye, 2022). Mes modifications vont à l'encontre de cette volonté de départ, sans pour autant alourdir de façon considérable la lecture des transcriptions. Le

[18] [o*] est utilisé pour des non-réalisations surprenantes, comme dans les exemples 260 et 261 p.252. Cette double fonction de l'annotation [x*] pourrait poser problème lors de la phase d'analyse, mais les cas incertains en sont exclus et ma thèse ne porte pas sur les consonnes finales habituellement non prononcées.

tableau comparatif 2.1 recense les éléments pris en compte dans les systèmes de codage utilisés pour les trois corpus pré-cités (Liégeois, 2014 : 183 ; Laks, 2008 : 241 ; Durand et al., 2002), ainsi que le système de codage que j'ai utilisé pour le corpus de tops 1 et le FJD.

	Corpus			
Annotation	**ALIPE**	**HPOL1**	**PFC**	**tops 1 & FJD**
Statut de la consonne de liaison	✓	✓		
(Non-)réalisation de la liaison	✓	✓	✓	✓(1)
(Non-)enchaînement de la liaison		✓	✓	
Présence d'une pause/hésitation ou d'un coup de glotte			✓	
Nature de la consonne de liaison attendue	✓	✓		✓(1)
Nature de la consonne de liaison réalisée	✓		✓	✓(1)
Part of speech (POS) du M1	✓	✓		✓(2)
POS du M2	✓	✓		
Lemme du M1				✓(2)
Nombre de syllabes du M1		✓	✓	
Genre du M1		✓		
Nombre du M1		✓		
Nasalité de la voyelle précédant une liaison en [n]			✓	

TABLEAU 2.1 : Tableau comparatif des éléments pris en compte par chaque système de codage de la liaison

Quelques précisions :

- pour ALIPE (Liégeois, 2014 : 183) :
  - les POS du M1 et du M2 constituent une unique annotation qui présente les valeurs suivantes : Déterminant - Nom, Pronom - Verbe, Expression figée, Verbe - Pronom, Adjectif - Nom, Nom pluriel - X, Avoir - X, Être - X, Verbe - X, Invariable - X, Cas Particuliers, Hors contexte
  - la nature de la consonne de liaison attendue peut être : /g, k, p, ʁ, t, n, z/
  - à cette liste, pour la nature de la consonne de liaison réalisée cette fois-ci, toute autre consonne peut être ajoutée (par exemple, /l/)
  - pour le statut de la consonne de liaison, il faut choisir entre les valeurs : catégorique, variable ou hors contexte de liaison
- pour HPOL1 (Laks, 2008 : 241, 256) :
  - la nature de la consonne de liaison attendue peut être : /z, t, r, p, n/
  - pour le statut de la consonne de liaison, il faut opter pour une des valeurs suivantes : obligatoire, facultative, erratique ou interdite
  - pour le genre du M1, le mot est soit masculin, soit féminin

  - pour le nombre du M1, le mot est soit singulier, soit pluriel
  - les POS du M1 ou du M2 peuvent être : Nom, Verbe, Déterminant, Adjectif, Pronom, Adverbe, Conjonction de coordination, Prep, Participe passé ou Mot outil
- pour le corpus PFC (Durand et al., 2002) :
  - pour le nombre de syllabes du M1, le mot peut être monosyllabique ou polysyllabique
  - pour la (non-)réalisation de la consonne de liaison : il peut y avoir réalisation, non réalisation ou incertitude quant à la réalisation
  - pour la nasalité de la voyelle précédant la consonne de liaison [n] : il peut s'agir d'une voyelle nasale ou d'une voyelle orale
- pour le corpus de tops 1 et le FJD :
  - remarque générale : l'annotation, afin de rester minimale, est effectuée en deux phases, indiquée à l'aide des numéros (1) et (2) dans le tableau 2.1
  - pour le statut de la consonne de liaison : cette information est absente car avec notre système de codage, nous évitons de catégoriser a priori les contextes de liaison (au contraire, le cadre théorique que j'adopte demande une catégorisation a posteriori des contextes de liaison, selon l'usage)
  - la nature de la consonne de liaison attendue est bien annotée, à l'exception toutefois des pataquès

Le tableau 2.1 nous présente ainsi des systèmes de codage de la liaison qui divergent. Notre système est l'un des moins complets, notamment lors de la première phase d'annotation, afin de rester minimal. Au contraire, l'annotation des corpus ALIPE, PFC et HPOL1 est plus complète mais aussi plus fastidieuse : les transcriptions manquent de lisibilité et le transcripteur est plus susceptible de faire des erreurs. Par exemple, lors d'une annotation des liaisons au fil de la transcription, oublier des liaisons invariables se produit plus rapidement qu'on ne le pense. J'en ai fait l'expérience au cours de l'annotation du corpus de tops 1. Une automatisation du repérage des contextes de liaison, comme nous allons l'aborder dans le paragraphe suivant, permet d'éviter que cela ne se produise.

Lors de la réunion susmentionnée du groupe de recherche s'intéressant à la liaison (2.1.3.2), nous avons réfléchi aux possibilités d'aide à l'annotation de ce phénomène grâce à l'outil informatique. En effet, la liaison avait jusqu'alors très souvent été annotée manuellement, comme dans le corpus PFC, au fur et à mesure de la transcription. Flora Badin, ingénieure d'études au Laboratoire Ligérien de Linguistique (LLL), avait

commencé à créer un outil sous Python (Badin, 2019) à la demande des chercheurs orléanais du groupe, afin de repérer automatiquement les contextes de liaison, ce qui empêche les oublis tout en permettant une annotation plus rapide. Cet outil transforme aujourd'hui des fichiers CLAN, transcrits et alignés avec le son, en documents ELAN enrichis. Le logiciel ELAN[19] (Brugman & Russel, 2004 ; Max Planck Institute for Psycholinguistics, 2022), qui permet l'alignement son/texte, l'annotation et la fouille de transcriptions linguistiques, a été développé par le Max Planck Institute for Psycholinguistics de Nijmegen, aux Pays-Bas, et est accessible depuis The Language Archive, une archive linguistique du même institut. Pour aboutir à des transcriptions ELAN,

> [t]out d'abord, les métadonnées ont été récupérées depuis les fichiers sources CLAN. Les transcriptions ont été converties au format .trs (logiciel Transcriber[, cf Barras et al., 1998]) grâce à l'outil TeiCorpo (cf [Modèles, Dynamiques, Corpus - UMR 7114 (MoDyCo), 2016] et Liégeois et al., 2015). Ce format nous a permis d'utiliser le logiciel Jtrans pour aligner notre transcription sur le signal sonore (voir [Laboratoire lorrain de recherche en informatique et ses applications - UMR 7503 (Loria), 2016 et] Cerisara et al., 2009) selon un découpage au niveau de séparateurs comme l'apostrophe, le tiret et l'espace sans distinction de frontière interne de mot. Ainsi nous obtenons aujourd'|hui mais aussi l'|arbre. Ce paramètre propre au logiciel n'est pas modifiable mais nous permet d'avoir un alignement optimal sur le son [...]. À partir du résultat obtenu, nous avons pu créer une transcription au format .eaf (logiciel ELAN) (Coutanson & Badin, 2021 : 138-139).

Il y a parfois eu des décalages entre le signal sonore et la transcription des paroles (pour le FJD, en raison d'éléments de langage non chantés et n'ayant de fait pas été transcrits et pour le corpus de tops 1, en raison de la présence d'un accompagnement instrumental). En outre, j'ai remarqué que certains tokens, présents dans l'énoncé, disparaissaient lors de la tokenisation, notamment pour les mots comprenant peu de caractères. C'est problématique lorsqu'il s'agit d'un M1 ou d'un M2, car le contexte de liaison n'est alors pas repéré comme tel et cela crée du silence. Néanmoins, ceci ne m'a semblé se produire que de façon très occasionnelle. Nos transcriptions ELAN sont dites enrichies car elles sont annotées en parties du discours et lemmatisées toujours à l'aide de l'outil TeiCorpo (Modèles, Dynamiques, Corpus - UMR 7114 (MoDyCo), 2016 ; Liégeois et al., 2015), qui recourt au logiciel Treetagger et à son jeu d'étiquettes pour le français (Schmid,

[19] ELAN est accessible à l'adresse suivante : https ://archive.mpi.nl/tla/elan.

2022; Stein, 2003; Stein & Schmid, 1995) et les liaisons sont pré-annotées automatiquement. Pour chaque annotation supplémentaire, une ligne d'annotation est ajoutée dans le fichier ELAN. Dans la figure 2.3, on voit :

- sur la première tire la transcription (avec un découpage à la pause pour le FJD),
- sur la seconde un alignement automatique au token,
- sur la troisième l'annotation de la liaison,
- sur la quatrième la lemmatisation,
- sur la cinquième l'annotation morphosyntaxique et
- à partir de la septième les métadonnées :
  - langue sur la septième tire,
  - nom du corpus sur la huitième,
  - code locuteur/chanteur sur la neuvième,
  - âge sur la dixième,
  - sexe sur la onzième,
  - rôle discursif sur la quatorzième,
  - profession sur la quinzième et
  - code postal de résidence sur la seizième.

L'outil récupère aussi toutes les annotations déjà présentes sur le fichier CLAN, telles les annotations phonétiques ou encore les commentaires (tires dépendantes %pho :, %exp : ou %com :). Ces éléments apparaissent sous forme de lignes d'annotation supplémentaires dans les fichiers ELAN. Pour le FJD, j'avais en outre demandé qu'une tire d'annotation de la modalité (parlé vs chanté) soit ajoutée (sixième tire), mais n'ayant transcrit que les parties chantées, je n'en ai pas eu l'utilité. D'autres enrichissements auraient pu être apportés. Par exemple, dans leur article de 2016, Greefhorst et van den Bosch (2016 : 20, 31), qui souhaitent prédire la réalisation des liaisons en corpus, utilisent d'autres pré-traitements intéressants, en renseignant la fréquence lexicale[20], le nombre de syllabes et la distance phonologique des mots 1 et 2, informations provenant de la base de données Lexique 3.8 (New et al., 2001). En reprenant les annotations manuelles apportées pour ALIPE, HPOL1 et PFC que je viens de lister, certaines semblent automatisables comme le repérage de pauses avant ou après la consonne de liaison. J'avais également mentionné en 1.2.2.3.1 l'intérêt d'une annotation en dépendances.

Intéressons-nous tout d'abord à la fonctionnalité de repérage des contextes de liaison proposée par notre outil. Pour cela, on doit fournir :

[20] La fréquence lexicale des mots 1 et 2 retenue provient d'un corpus de sous-titres de films dans Lexique 3 (New & Pallier, [s. d.] : 10 ; Greefhorst & van den Bosch, 2016 : 20).

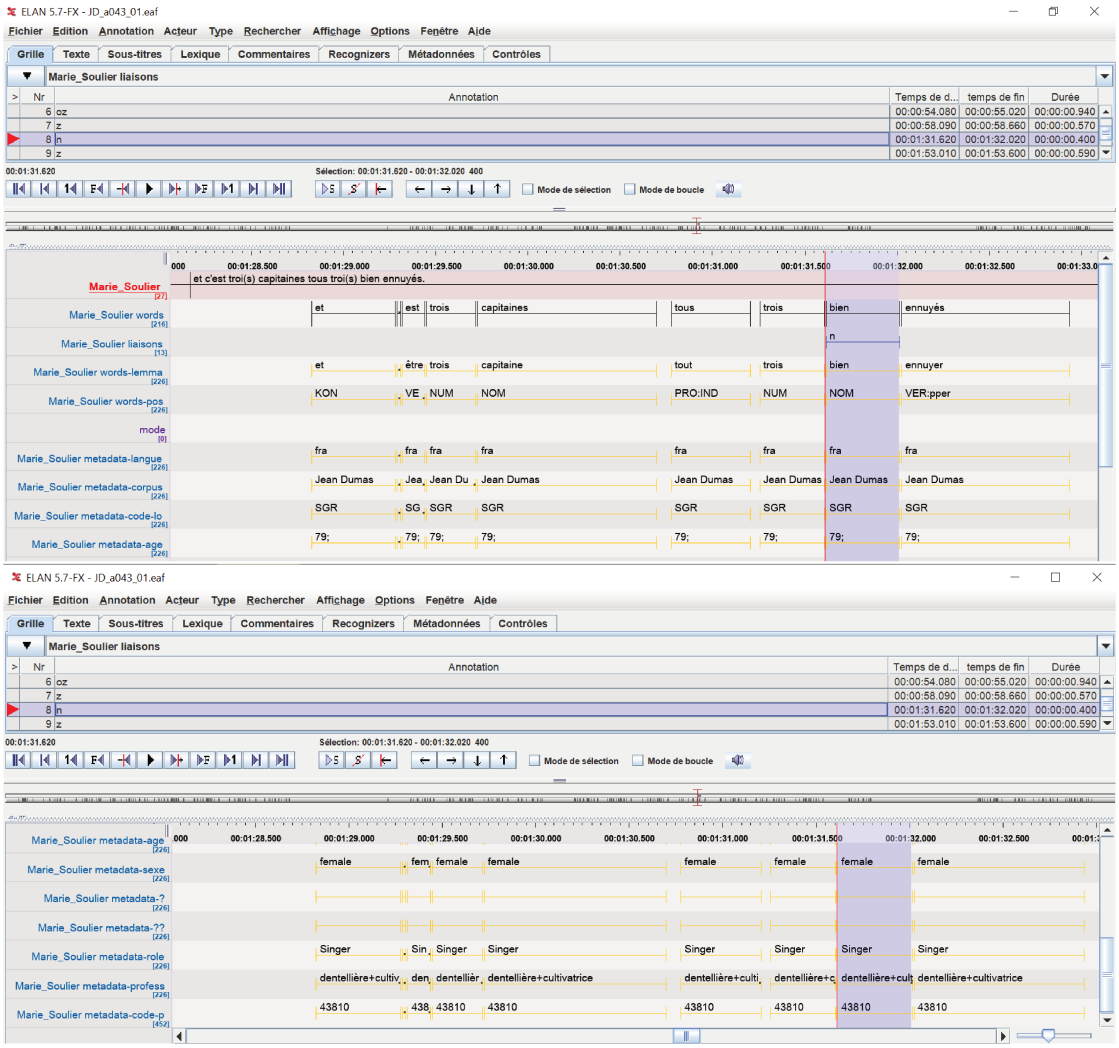


FIGURE 2.3 : Exemple de fichier ELAN enrichi (JD_a043_01)

- une liste de derniers caractères de M1 liaisonnants : pour mes corpus ‹d, g[21], m[22], n, p, r, s, t, x, z›.
- une liste de premiers caractères de M2 liaisonnants : pour mes corpus ‹a, A, à, À, â, Â, e, E, è, é, É, ê, Ê, F, h, H, i, I, î, Î, L, M, N, o, O, R, S, u, U, w, W, X, y, Y›. Certains graphèmes consonantiques majuscules sont présentés dans cette liste par sécurité, afin d'éviter de passer à côté des sigles liaisonnés débutant par une

[21] Pour des liaisons en [k].

[22] J'avais conservé ‹m› par sécurité pour les nasales finales prenant le graphème ‹m›, ce qui ne s'est pas révélé utile en fin de compte.

lettre à initiale vocalique (ex : son [on] HLM ; Tops 1, Jean Ferrat, La Montagne, 1965 [1964]). Pour limiter le bruit, celles-ci pourraient être retirées. Il est difficile d'atteindre une juste mesure entre bruit et silence : par exemple, les consonnes graphiques épelées hors sigle échappent à mon investigation (voir exemple 75 p.139), à moins qu'elles n'aient été annotées au fil de la transcription.

- voire une liste d'exclusion pour les M1. Par exemple, j'ai fourni la liste suivante pour le corpus de tops 1 (les mots sont issus d'un calcul de fréquence sous CLAN) : .*ar, .*ir, .*or et .*ur (sauf *monsieur*), qui correspondent à l'ensemble des mots finissant par les digrammes ‹ar›, ‹ir›, ‹or› et ‹ur› ; *adios*, *Alison*, *août*, *banana + split*, *Barbès*, *basket*, *beat*, *blues*, *Boris*, *Boston*, *business*, *cancer*, *card*, *cher*, *clebs*, *correct*, *Dallas*, *Delegation*, *direct*, *enfer*, *ex*, *fabulous*, *fan*, *feeling*, *fer*, *film*, *flirt*, *foot*, *footing*, *from*, *Gims*, *hélas*, *hier*, *hip + hop*, *hiver*, *hop*, *hum*, *index*, *internet*, *jadis*, *jazz*, *jean*, *jump*, *Jupiter*, *Khaled*, *Kleenex*, *lips*, *Lucifer*, *man*, *net*, *number*, *Mahmoud*, *Maureen*, *mer*, *Nolwenn*, *ouest*, *pioneer*, *post + it*, *prout*, *rap*, *Regglyss*, *revolver*, *rhum*, *Rothen*, *Saddam*, *scooter*, *sept*, *set*, *shopping*, *short*, *slim*, *sud*, *summum*, *swing*, *tennis*, *Texas*, *test*, *t + shirt*, *twist*, *Vénus*, *Weston*, *Wiltord*, *yep*, *yes*, *zut*.

Nous ne sommes pas les seuls à avoir tenté d'automatiser un minimum le repérage des contextes de liaison et l'annotation de celles-ci : Adda-Decker et al. (2012) l'avaient déjà fait pour le corpus NCCFr. Nous n'avions pas connaissance de la méthodologie mise en place par ce groupe de recherche lors de l'établissement de notre propre méthodologie. Afin de repérer automatiquement les contextes de liaison, Adda-Decker et al. (2012 : 117) ont également commencé par délimiter les M1 potentiellement liaisonnants en fonction de leur graphème final (-‹d›, -‹n›, -‹p›, -‹r›, -‹s›, -‹t›, -‹x›, -‹z›), avant d'éliminer les mots pour lesquels ce graphème n'est pas une consonne de liaison, que ce soit parce qu'il s'agit de consonnes fixes, mais pas seulement : ces chercheurs ont aussi rejeté les noms propres ou encore les noms finissant par une nasale. Ensuite, ces auteurs ont voulu délimiter les M2 liaisonnants en ne conservant que ceux commençant par une voyelle ou une glissante, mais ont fait le choix d'exclure les mots débutant par des ‹h›- ou des glissantes disjonctifs, par une hésitation ou une répétition (Adda-Decker et al., 2012 : 117).

Pour ma part, j'ai une conception très large de ce qu'est un contexte de liaison. Pour moi, la frontière de mots entre tout M1 finissant par les caractères consonantiques finaux listés ci-dessus, muets en isolation, et tout M2 débutant par les caractères vocaliques initiaux listés ci-dessus (pour rappel, les graphèmes consonantiques n'ont été ajoutés que pour les cas où le nom de la lettre, à initiale vocalique, est prononcé), qu'il s'agisse d'une voyelle ou d'une glissante (les réalisations *des yeux* [dezjø], *des oiseaux*

[dezwazo] et *des huîtres* [dezɥitʁ] ont été mentionnées en 1.2.1), constitue un site potentiel de liaison et ce, « quel que soit le contexte prosodique, syntaxique de l'énoncé, le statut lexical ou encore morphosyntaxique des mots 1 et 2 » (Coutanson & Badin, 2021 : 136). Je défends cette conception très large du contexte de liaison pour deux raisons : par précaution lors de l'étude de nouveaux corpus et par souci de comparabilité. En ce qui concerne la première raison, la question s'est surtout posée pour moi lors de l'étude du FJD qui présente des réalisations inhabituelles, à des frontières de mots qui ne sont pas considérées comme des contextes de liaison en français contemporain (voir par exemple le cas des locutions prépositives *par/pour rapport (à)* en 5.2.3, qui peuvent être liaisonnées dans le fonds). Seuls des projets tels qu'InterPhonologie du français contemporain (IPFC), le volet FLE/FLS du corpus PFC, dans lequel les chercheurs se devaient d'être très prudents, ont une conception aussi large que la mienne. C'est en effet dans les variétés les moins standard, que ce soit parce qu'il s'agit de productions d'apprenants du français, de données acquisitionnelles, diachroniques ou encore régionales, qu'il existe un risque : si l'annotation est faite au fil de la transcription, les réalisations atypiques devraient correctement être annotées mais les contextes similaires ont de grandes chances de ne pas être annotés comme sites potentiels de liaison. Pour ce qui est de la deuxième raison invoquée, la délimitation des contextes de liaison est extrêmement fluctuante selon les linguistes, ce qui ne permet aucune comparaison réelle entre les résultats des différentes études et constitue un frein important au cumul des connaissances scientifiques. Au minimum, chaque étude devrait présenter très précisément ce qui a été considéré comme site potentiel de liaison, tels Dugua et Baude (2017 : 44-45) qui indiquent ne pas prendre en compte les contextes nom singulier + X, nom propre + X, X + *oui* et X + *ouais* comme des contextes de liaison possibles, ainsi que deux mots appartenant à des groupes syntaxiques différents (comme *on ne sait pas on manque d'information*), prononcés par des locuteurs différents ou séparés par une pause. Le partage systématique des corpus permettrait également une reproductibilité des études par d'autres chercheurs.

Suite à l'automatisation du repérage des contextes de liaison, il faut se demander dans quelle mesure l'automatisation de l'annotation des liaisons est possible. L'outil conçu par Flora Badin pré-annote automatiquement les liaisons du corpus en fonction des sites potentiels repérés. Par défaut, les liaisons du FJD sont pré-annotées comme réalisées et les liaisons du corpus de tops 1 comme non réalisées, selon la tendance générale du corpus. L'outil détermine la nature de chaque consonne de liaison en fonction du dernier caractère du M1. Un retour manuel est fait pour chaque contexte, grâce au balayage de contexte en contexte permis par ELAN (onglet *Grille* puis sélection de la

tire *liaisons* pour le chanteur), permettant une annotation « à la chaîne » des liaisons, avec une réécoute immédiate (voir figure 2.3). Si un contexte de liaison n'en est pas un, il est supprimé, puis à l'aide d'un menu déroulant, la pré-annotation est manuellement corrigée : selon notre système de codage mentionné en 2.1.3.4, on peut choisir entre réalisation du segment (z, n, t, r, p, k), non-réalisation du segment (oz, on, ot, or, op, ok), pataquès (*z, *n, *t, *r, *p, *k) ou hésitation/consonne finale habituellement non prononcée (z*, n*, t*, r*, p*, k*). Les corrections apportées portent principalement sur le caractère réalisé/non réalisé du segment ou accessoirement sur la nature du segment en cas de pataquès. Les chants à l'unisson posent problème, comme le notaient déjà Nardy et al. (2014 : 251) dans un corpus d'enfantines : les cas ambigus ont été codés comme incertains (Coutanson & Badin, 2021 : 140). Il est important de préciser que ces annotations n'ont pas été vérifiées par d'autres annotateurs, contrairement à un principe fondamental de la linguistique de corpus, ce qui aurait permis de calculer un accord inter-annotateurs.

Afin d'automatiser l'annotation des liaisons, Adda-Decker et al. (2012 : 118) sont allés jusqu'à la mise en place d'un alignement automatique entre le signal sonore et la transcription, au mot comme nous l'avons fait, mais aussi au phonème. Un segment supplémentaire en frontière de mots dans le signal sonore est considéré comme une liaison réalisée si celle-ci est indiquée dans un dictionnaire de prononciation de référence. Les chercheurs ont obtenu de bons résultats, avec un taux d'erreurs inférieur à 5% (2012 : 119). Notons toutefois que cette technique, fondée sur l'unique présence/absence d'une consonne supplémentaire, ne permettrait pas l'annotation de pataquès. Dans la récente étude d'Hutin et al. (2022), « le système dispose [...] de modèles de phones pour chaque phonème du français, qui lui permettent d'identifier les phones sur un spectrogramme ». Ainsi outre la présence/absence de liaisons, leur nature pourrait être annotée. Le dictionnaire est par ailleurs enrichi pour autoriser /z, n, t, p/ dans n'importe quel contexte de liaison, mais pas //ʁ//, en raison de ses nombreuses variantes en français, ce qui permettrait cette fois-ci de détecter les pataquès, sauf ceux en //ʁ//. En étudiant la séquence verbe *être* conjugué + *un/une* (déterminant), ces chercheurs notent environ 7% d'erreurs d'alignement, ce qui semble comparable aux 5% d'Adda-Decker et al. (2012). Toutefois, il s'avère impossible de le vérifier car ces derniers ne communiquent pas leurs effectifs. De surcroît, Hutin et al. (2022) notent environ 14% d'inéquation entre consonne graphique finale et consonne détectée automatiquement, alors que la réalisation de pataquès dans ces contextes est minime (8/1659 occurrences). Cela est dû à des causes techniques (bande passante insuffisante, audio inaudible) ou à des causes liées à la situation discursive (bruit de fond, disfluences, chevauchements, hypo-articulation,

autre modalité, intensité atypique). De la sorte, les propositions d'Hutin et al. (2022) étaient prometteuses pour l'étude du pataquès, mais le taux d'erreurs de repérage de liaisons/pataquès est bien supérieur aux pataquès correctement annotés.

Pour conclure, j'ai présenté le système de codage adopté par une équipe de chercheurs acquisitionnistes travaillant sur la liaison, système que j'ai adapté pour les besoins de cette thèse sans manquer de le comparer aux systèmes de codage existants pour la liaison (corpus ALIPE, HPOL1 et PFC). J'ai précisé que seuls les sandhis externes inattendus hors contexte de liaison doivent être annotés dès l'étape de transcription sur CLAN. J'ai par la suite présenté un outil d'annotation des liaisons et d'enrichissement des transcriptions (étiquetage morphosyntaxique, lemmatisation, alignement automatique au mot), que j'ai utilisé, outil développé par Flora Badin. D'autres enrichissements pourraient encore être envisagés : annotation automatique de la fréquence lexicale, du nombre de syllabes et de la distance phonologique des mots 1 et 2, présence d'une pause, annotation en dépendances… Pour repérer les sites potentiels de liaison, notre outil a besoin d'une liste de derniers caractères de M1, d'une liste de premiers caractères de M2, voire d'une liste d'exclusion pour les M1. D'autres chercheurs font preuve de moins de prudence : le graphème final ‹g›, les noms propres et les noms à nasale finale en M1, les glissantes et les ‹h› disjonctifs à l'initiale de M2 ne sont pas considérés comme liaisonnants pour Adda-Decker et al. (2012). Cela m'a amenée à proposer une réflexion plus personnelle sur ma conception de ce qu'est un contexte de liaison, une conception large mais plus prudente lors de l'étude de nouveaux corpus et permettant une meilleure comparabilité inter-corpus. Pour continuer, je suis revenue sur la préannotation automatique des liaisons par l'outil et sur leur correction manuelle, sans l'aide d'autres annotateurs. C'est surtout en ce qui concerne l'annotation automatique de la liaison que la méthodologie développée par Adda-Decker et al. (2012) s'est révélée innovante. Les corrections apportées par Hutin et al. (2022) sont très intéressantes pour l'annotation du pataquès, mais le taux d'erreurs concernant les liaisons et les pataquès demeure trop élevé.

2.1.3.5 Réécoute et vérification

La cinquième étape comprend une vérification du fichier .cha produit, au fil d'une réécoute, en apportant au besoin une correction à certains éléments et en balisant de façon plus adéquate certains passages. J'ai en fait effectué deux vérifications : une à la fin de la transcription/annotation de chaque chanson et une seconde, pour chaque

chanson, une fois le corpus terminé afin d'uniformiser la transcription/annotation sur des points spécifiques et de corriger manuellement des occurrences qui n'auraient pas été prises en compte et été modifiées lors de requêtes d'uniformisation.

Cette dernière étape consiste donc, en réécoutant l'audio, à vérifier la transcription, l'annotation et l'alignement son/texte, juste après la transcription du document et à nouveau une fois le corpus terminé.

Ainsi, selon la méthodologie que j'ai mise en place, pour constituer de nouveaux corpus, que les données soient compilées ou recueillies, il faut tout d'abord obtenir un fichier son et établir un plan de nommage, puis préparer un fichier .cha à associer à ce fichier son en renseignant les métadonnées de l'enregistrement, avant d'aligner le son avec des portions de transcription. Le travail le plus conséquent est par la suite de transcrire et d'annoter la production linguistique en associant chaque tire à un participant, avant de vérifier le tout lors de deux réécoutes de vérification. Pour terminer, un outil développé par Flora Badin permet un dernier enrichissement des données et notamment le repérage automatique de contextes de liaison et la pré-annotation automatique des liaisons du corpus, qui sont ensuite vérifiés manuellement. Revenons maintenant à la méthodologie générale : il est temps d'analyser les données.

2.1.4 Analyse fine des corpus les plus pertinents

Après avoir navigué entre état de l'art et collection d'exemples attestés, après avoir choisi des corpus et les avoir carottés, après avoir constitué des corpus faisant auparavant défaut (qu'on peut carotter en cas de doute sur leur pertinence), la méthodologie établie dans le cadre de cette thèse propose de se plonger dans les données. Comme je l'expliquerai en 4.1, deux corpus sont à analyser plus finement : le FJD et le corpus de tops 1.

Les corpus transcrits et annotés servent en premier lieu de réservoirs d'exemples (Leech, 1991 : 13) pour des analyses qualitatives, comme on peut le voir au fil de ce manuscrit. En second lieu, le choix d'ELAN pour la dernière phase d'annotation des données permet une exportation facile des éléments qui nous intéressent vers un tableur (Coutanson & Badin, 2021 : 140) afin d'effectuer des analyses quantitatives, grâce à la fonction Rechercher/Recherche structurée de plusieurs eaf…/*Multiple Layer Search*. Pour mener à bien ces analyses, j'ai extrait les informations suivantes depuis ELAN pour chaque contexte de liaison :

- *TranscriptionName* : nom de la chanson concernée, selon les plans de nommages présentés en 2.2.1 et 2.3.1
- *BeginTime* : balise initiale de la tire principale dont il est question
- *TierParticipant* : nom du ou des chanteur(s) impliqué(s)
- *Annotation1-1* : annotation de la liaison ou du pataquès
- *Annotation2-1* : M1
- *Annotation2-2* : M2
- *Annotation3-1* : POS du M1
- *Annotation3-2* : POS du M2
- *Annotation4-1* : lemme du M1
- *Annotation4-2* : lemme du M2

ELAN permet aussi d'extraire d'autres informations comme les métadonnées, mais moins de contextes peuvent alors être extraits.

Pour des analyses qualitatives, les corpus tiennent donc lieu de réservoirs d'exemples. Pour des analyses quantitatives, ELAN nous permet d'exporter les transcriptions et annotations liées à chaque contexte de liaison vers un tableur. Je vais continuer en reprenant ma chaîne de traitement pour les deux corpus que j'ai constitués, en mettant en avant les caractéristiques de cette chaîne de traitement pour chaque corpus, en commençant par la compilation du corpus de tops 1.

2.2 Traitement du corpus de tops 1

2.2.1 Téléchargement du fichier son et mise en place d'un plan de nommage

Je vais commencer par expliquer comment j'ai sélectionné les chansons de ce premier corpus, puis comment j'ai obtenu le fichier son de chaque chanson, avant de présenter le plan de nommage que j'ai ici mis en place.

En premier lieu, j'ai recherché des chansons ayant accédé à la première place de classements musicaux réalisés pour la France. Je n'ai conservé que les chansons d'expression française ou en partie d'expression française, car je m'intéressais aux phénomènes de liaison et potentiellement de pataquès dans ces titres, phénomènes absents dans les autres langues ayant accédé au top 1. Pour ce faire, j'ai consulté rapidement les paroles des chansons classées n°1 (je présenterai en 2.2.3 comment j'ai accédé aux paroles des chansons) et si je relevais un passage en français (un vers minimum), la

chanson était sélectionnée pour le corpus constitué.

J'ai tout d'abord eu l'idée de compiler un corpus débutant en 1984, date à laquelle a débuté le Top Singles[23] (Lesueur & Durand, 2022a), comprenant tous les titres ayant accédé à la première place de ce classement musical. Davantage connu sous le nom de Top 50, car il mettait en avant au départ cinquante titres hebdomadaires, ce classement était réalisé par deux instituts de sondage, Nielsen et Ipsos (Lesueur & Durand, 2022a). Ce classement aurait pris fin en 1993, remplacé en 1995 par un Top Singles commandité par le SNEP et réalisé par l'Institut français d'opinion publique (IFOP) (Lesueur & Durand, 2022a). Au départ, le Top Singles prenait en compte uniquement les singles physiques, puis également les téléchargements payants et le streaming dès 2016, avant de se limiter au streaming payant dès 2018[24] (SNEP, 2022a). De cette façon, il ne comptabilise « ni la diffusion dans les médias, ni le prêt entre particuliers ou en bibliothèque, ni la copie, ni le téléchargement illégal ou encore la revente » (Coutanson & Badin, 2021 : 136). Je collectais les informations concernant le Top Singles sur *Pure charts* (http ://www.chartsinfrance.net/), un site fondé par des passionnés de musique s'intéressant aux ventes de singles et d'albums en France, publiant des actualités musicales et comprenant un forum, site racheté en 2011 par le groupe Webedia (Pure charts, 2022). J'avais contacté l'IFOP à propos d'un bref laps de temps où le classement n'était pas tout à fait clair sur Pure charts. Je n'ai pas obtenu de réponse. Depuis la semaine du 11 décembre 2020, le classement ne semble plus actualisé sur le site. À partir de la semaine du 18 mai 2018 (au 04.09.2022), le classement est aussi disponible directement sur le site du SNEP (https ://snepmusique.com/). Une difficulté s'est présentée : pour la période où le classement est archivé sur les deux sites, des discordances sont visibles... En partant d'un site non producteur des données, il faut s'attendre à des erreurs dans l'archivage des titres numéro 1 pour la France et ainsi dans le corpus que j'ai constitué. Pour terminer, concernant la date de fin du corpus, l'année 2017 a fait office de clôture, car je l'ai compilé courant 2018. En fonction de la date de début d'un second sous-corpus que je vais à présent mentionner, cela me permettait également de subdi-

[23] Selon les règles du Syndicat national de l'édition phonographique (SNEP), « [s]ont considérés comme singles les phonogrammes contenant de 1 à 4 titres différents » : on parle de *format court* (Syndicat national de l'édition phonographique, 2021).

[24] Les ventes physiques des grandes surfaces alimentaires, des grandes surfaces spécialisées, des disquaires indépendants, du e-commerce et des sites d'artistes ou de leurs labels sont comptabilisées, hormis les produits exclusifs (voir SNEP, 2022b pour plus de précisions). Les plateformes de streaming prises en compte sont les suivantes : Apple Music, Spotify, Deezer, Qobuz, Napster et Google Music, pour des titres écoutés au moins trente secondes, et 150 streams équivalent à une vente physique ou à une vente en téléchargement (SNEP, 2022b).

viser le corpus total en décennies afin de réaliser des analyses micro-diachroniques.

Par la suite, il m'a semblé qu'il serait plus intéressant de travailler sur une période plus étendue. J'ai donc regroupé les titres entre 1984 et 2017 dans un sous-corpus et j'ai décidé de constituer un deuxième sous-corpus antérieur à cette période, ce qui n'allait pas de soi étant donné que le Top Singles n'existait pas avant 1984. La page Wikipédia *Liste des singles numéro un en France* (Wikipédia, 2022) faisait mention du site *Infodisc* (https ://infodisc.fr/), qui propose un classement des titres depuis la semaine du 5 octobre 1955[25] (Lesueur & Durand, 2022b). J'ai choisi de faire débuter ce deuxième sous-corpus en 1956 afin d'avoir une année complète : le corpus total couvre ainsi une période allant de 1956 à 2017. Le site Infodisc propose une synthèse hebdomadaire de différents hit-parades, tentant de dépasser les défauts de certains classements musicaux, comme la non-prise en compte de certains tubes n'ayant pas été diffusés au format single (mais uniquement au sein d'albums) ou la prise en compte des singles dans leur entièreté plutôt que des diverses chansons les constituant, faussant les classements quand plusieurs chansons d'un single deviennent des tubes (Lesueur & Durand, 2022c). Le classement proposé sur ce site n'est pas fonction des ventes de singles mais synthétise les informations disponibles d'une époque donnée (Lesueur & Durand, 2022a). Lorsque j'ai constitué ce sous-corpus en 2018, les informations référencées sur la page Wikipédia n'étaient pas accessibles sur Infodisc. Ce dernier était monétisé et j'ai alors pris un abonnement d'un mois (que j'ai annulé par la suite), afin de vérifier les informations présentées sur Wikipédia. Je n'ai pas davantage trouvé les références concernées, et j'ai finalement utilisé Wikipédia comme source afin de sélectionner les chansons de ce deuxième sous-corpus. De cette façon, les chansons sélectionnées pour ce sous-corpus ne sont pas toujours en accord avec les données actuellement présentées sur Infodisc, qui ont par la suite été mises à jour (fin 2018 !) et qui sont aujourd'hui accessibles de manière non monétisée.

Une fois cette sélection effectuée, il m'a fallu accéder aux fichiers audio de ces chansons, que j'ai récupérés un à un depuis Youtube (https ://www.youtube.com/). Il m'a semblé que c'était une des façons les plus rapides pour obtenir les fichiers, sans toutefois garantir la meilleure qualité de son.

En ce qui concerne le plan de nommage de chaque chanson du corpus, j'ai décidé d'adopter le format suivant : (année de la première accession au top 1)_(premier jour de la semaine-mois de la première accession au top 1)_(dernier jour de la semaine-mois

[25] Accessible dans l'onglet *Les chansons/Les Chansons N° 1* (colonne de droite)/*5 0's*, *6 0's*, *7 0's*, *8 0's*, *9 0's*, *2000's* et *2010's* (colonne de droite).

de la première accession au top 1)_(titre de la chanson sans caractères spéciaux[26]). Ce qui donne par exemple :

- 1968_13-07_19-07_Petite_fille_de_francais_moyen (Tops 1, Sheila, *Petite fille de français moyen*, 1968 [1968])
- 1987_03-04_09-04_Viens_boire_un_p-tit_coup_a_la_maison (Tops 1, Licence IV, *Viens boire un p'tit coup à la maison*, 1987 [1986])
- 2010_03-12_09-12_Oui_mais_non (Tops 1, Mylène Farmer, *Oui mais... non*, 2010 [2010])
- 2016_22-07_28-07_ELEPHANT (Tops 1, Booba, *É.L.É.P.H.A.N.T*, 2016 [2016])

Placer l'année en premier permettait généralement de lister les titres par ordre d'accession au top 1, même si mettre le mois en deuxième position et le jour en troisième position aurait été préférable dans la même optique.

Pour résumer, j'ai constitué un corpus de chansons (en partie) d'expression française ayant été classées premières pour la France entre 1956 et 1984 pour un premier sous-corpus et entre 1984 et 2017 pour un second sous-corpus. Nous avons vu que la pertinence des titres sélectionnés est relative. En effet, le classement du Top Singles n'a pas été archivé en ligne par les instituts de sondage qui l'ont réalisé mais seulement par des amateurs de musique ou professionnels du secteur de la musique. Avant 1984, c'est encore plus délicat car les titres proposés relèvent d'une synthèse des tendances de l'époque. Par ailleurs, les classements sont déconnectés des moyens d'accès hors achat à la musique enregistrée ou en retard sur ceux-ci. Nous avons aussi vu que ces classements comportaient des failles et qu'ayant moi-même finalisé un corpus avec les références obtenues courant 2018, je ne suis pas revenue sur les titres sélectionnés par la suite afin de les actualiser. Ce corpus pourrait de la sorte être grandement amélioré dans sa composition, bien qu'il reste tout de même assez représentatif d'une époque. S'il devait l'être, il faudrait attendre qu'il y ait un peu plus de consensus en ce qui concerne les titres consacrés au sein des différents classements. Enfin, les audios de ce corpus ont été récupérés depuis Youtube et j'ai mis en place un plan de nommage indiquant la date d'accession au top 1 suivie du titre de la chanson. L'étape suivante consistait à associer un fichier .cha à chaque fichier son.

[26]Les accents, cédilles, points ou encore points de suspension ont été retirés et les apostrophes ont été transformées en traits d'union.

2.2.2 Préparation d'un fichier .cha

Dans cette étape, je créais des fichiers .cha suivant le même plan de nommage que celui vu en 2.2.1. Notre méthodologie générale (voir 2.1.3.2) indiquait qu'il fallait renseigner au cours de cette étape les métadonnées suivantes : âge (en années uniquement), sexe présumé et rôle du chanteur, langue de la chanson et nom de la transcriptrice. Pour le corpus de tops 1, j'ajoutais aussi les métadonnées suivantes : prénom et nom du chanteur ou son nom de scène (par exemple Dalida et non Iolanda Gigliotti) ainsi que son lieu de naissance (ce dernier est indiqué sous la forme d'un code postal[27] ou d'un nom de pays pour les chanteurs nés à l'étranger). L'âge du chanteur était obtenu en fonction de sa date de naissance. Sa date et son lieu de naissance étaient le plus souvent récupérés sur Wikipédia (https ://fr.wikipedia.org/), ce qui n'est pas sans risque d'erreur. J'aurais pu indiquer le lieu de vie du chanteur au moment de la sortie de la chanson plutôt que son lieu de naissance, mais ce dernier était plus facile à obtenir et me semblait plus pertinent. Pour ce corpus, seul le rôle Singer est utilisé. Lorsqu'une chanson présente plusieurs chanteurs, les codes participants étaient numérotés en SGR1/SGR2 et ainsi de suite.

Ainsi, chaque fichier .cha créé porte le même nom que l'audio de la chanson, et les métadonnées suivantes lui ont été associées : prénom et nom ou nom de scène, âge, sexe présumé, lieu de naissance et rôle du chanteur ; langue de la chanson et nom de la transcriptrice, sans oublier le fichier audio à associer. Dans la continuité, le traitement suivant consistait à aligner le son avec le texte après avoir obtenu les paroles de chaque chanson.

2.2.3 Obtention des paroles et alignement son/texte

Pour le corpus de tops 1, la méthodologie générale est légèrement adaptée au cours de cette étape. J'exposerai un moyen que j'ai utilisé afin de transcrire ce corpus plus rapidement, avant de revenir sur l'objectif principal de cette étape, soit le balisage du fichier .cha comme vu en 2.1.3.3.

Les chansons présentent une caractéristique qui facilite leur transcription : leurs paroles sont parfois copiées dans un objectif de transmission. Pour les chansons issues de l'industrie du disque, les paroles peuvent être fournies sur les livrets de CD, sur des logiciels de karaoké, sur des sites internet spécialisés ou encore sur des plateformes de

[27] Voir note de bas de p.30.

streaming. Pour éviter de perdre du temps, j'ai décidé de récupérer les paroles essentiellement sur le site Lyrics Translate (https ://lyricstranslate.com/), un site collaboratif de dépôt et surtout de traduction de paroles de chansons, et dans une moindre mesure sur le site Paroles.net (https ://www.paroles.net/), afin de transcrire plus rapidement les chansons. Les paroles n'ont pas été aspirées sur le web mais recherchées manuellement puis copiées sur le fichier .cha. Je reviendrai plus en détail sur la transcription de ce corpus dans le point suivant qui lui est consacré (2.2.4).

Pour le corpus de tops 1, j'ai conservé le découpage en vers qu'on trouve habituellement dans les paroles de chansons. Je n'avais plus qu'à aligner chaque vers avec le fichier sonore en renseignant pour chacun d'entre eux le code du chanteur associé.

Retenons donc que je me suis servie de paroles mises en ligne afin d'écourter mon temps de transcription, paroles que j'ai copiées sur mon fichier .cha avant de réaliser l'alignement son/texte. Il était alors temps de se lancer dans l'étape la plus conséquente, soit la transcription et l'annotation du corpus.

2.2.4 Transcription et annotation du document sonore

Dans cette partie, je reviendrai sur la transcription et l'annotation du corpus de tops 1 à l'aide de CLAN. Je reviendrai sur les modifications des paroles obtenues en ligne, puis je présenterai plus spécifiquement deux annotations employées pour ce corpus.

Les paroles recueillies en ligne ont subi de nombreuses modifications afin qu'elles se conforment aux conventions de transcription et d'annotation de CLAN (voir 2.1.3.4). Elles ont aussi été rigoureusement revues, les entités nommées et les termes argotiques ont été vérifiés, les interjections propres au chant ont été dénombrées avec précision et les vers bissés ont été doublés.

En ce qui concerne l'annotation de ces chansons issues de l'industrie du disque et plus particulièrement des schwas, les non-réalisations de ‹e› graphiques sont indiquées entre parenthèses, à l'exception des schwas en fin de mot qui n'ont pas été annotés (cf *donne* dans l'exemple 69 p.138). En effet, ceux-ci ne sont habituellement pas produits en français standard contemporain. Loïc Liégeois, pour l'annotation des schwas de son corpus de thèse (ALIPE ; voir Liégeois, 2014), avait opté pour cette convention.

Pour ce qui est de l'annotation des liaisons dans le corpus des tops 1, j'avais indiqué en 2.1.3.4 que seuls les sandhis inattendus hors contexte de liaison devaient impérativement être annotés lors de la présente étape. Pour ce premier corpus, j'ai pourtant décidé d'effectuer une annotation au fil de la transcription, ce qui ne me semblait pas

contre-productif, la tâche étant très rapide lors de l'écoute d'un vers.

J'ai mentionné ici certains choix qui me sont propres lors de la transcription et de l'annotation du corpus de tops 1 (ne pas annoter les schwas en fin de mot, annotation complète des liaisons pour ce premier corpus). Il ne restait alors plus qu'à vérifier les transcriptions/annotations en réécoutant une dernière fois les audios.

2.2.5 Réécoute et vérification

Dans cette dernière étape, les fichiers .cha étaient vérifiés au cours d'une réécoute des audios. La constitution de ce corpus aura nécessité entre deux et trois mois de travail exclusif et intense sur ce projet, sans compter cette dernière étape de vérification (double pour rappel).

À présent, synthétisons brièvement la méthodologie qui a été mise en place pour le traitement du corpus de tops 1, subdivisé en deux sous-corpus, 1956-1984 et 1984-2017. Pour commencer, j'ai présenté en détail comment les chansons avaient été sélectionnées. Cette sélection a été semée d'embûches, et on ne peut pas affirmer qu'elle reflète de façon incontestable les titres ayant accédé à la première place de classements musicaux pour la France. Les audios ont été obtenus via Youtube avant d'être renommés, puis ils ont été associés à un fichier .cha. Les métadonnées de ce dernier étaient par la suite renseignées. Pour continuer, je récupérais les paroles des chansons en ligne et je conservais leur subdivision en vers avant d'aligner le son avec ces paroles, tout en associant chaque tire à un chanteur. Je vérifiais ensuite les paroles et je les adaptais aux conventions de transcription et d'annotation de CLAN. J'ai présenté deux annotations (schwas et liaisons) propres à ce corpus. Pour finir, je contrôlais les fichiers .cha produits. Penchons-nous maintenant sur le traitement du FJD.

2.3 Traitement du fonds Jean Dumas

J'ai réutilisé pour cette thèse une grande partie des quatre premiers corpus du FJD, mis à disposition des internautes sur le site de la BIPO (cf tableau 3.1, p.184). Pour chaque chanson, on a généralement accès à :

- un audio au format .mp3[28], qui constitue une performance unique d'un chanteur,

[28]Dans le *Guide des bonnes pratiques pour le traitement documentaire des archives sonores inédites*, il est conseillé de numériser les audios sans compression, aux formats WAV ou BWF pour la conservation, puis

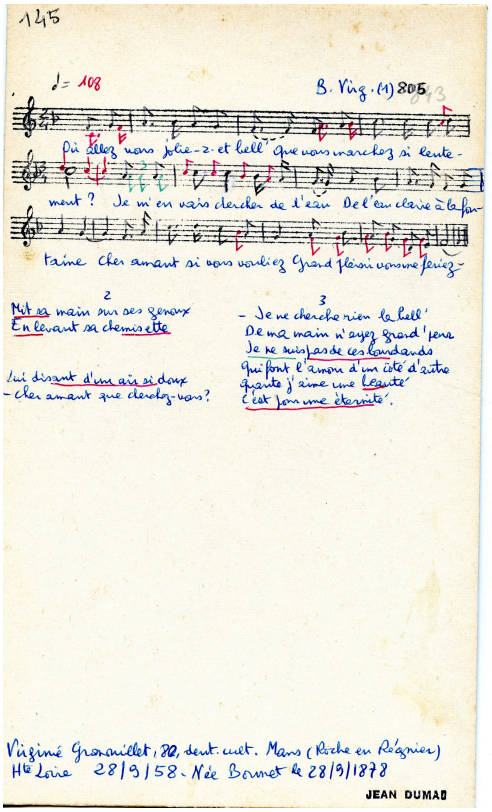
parfois plusieurs s'ils ont miraculeusement été conservés par Jean Dumas (il peut y avoir eu deux versions quand le chanteur propose à nouveau une œuvre déjà interprétée lors d'une précédente séance de travail ou en cas de première prise ratée ; l'absence de fichier sonore, fait rare, exclut automatiquement la chanson de mon analyse),
- une fiche manuscrite rédigée par Jean Dumas, téléchargeable au format .pdf (cf exemple 2.4a ; très rarement cette fiche est absente), voire
- une fiche supplémentaire tapuscrite au format .pdf (cf exemple 2.4b), comprenant également paroles et mélodie. Ces fiches ont été rédigées exclusivement pour le corpus centré sur le répertoire de la chanteuse Virginie Granouillet, sûrement lors de la mise en ligne du corpus. Les transcripteurs semblent être repartis des données primaires car ces fiches peuvent différer des transcriptions de Jean Dumas.

Dans ma thèse, j'ai ainsi réutilisé une grande partie des quatre premiers corpus de chansons de tradition orale du FJD. Pour chaque chanson, on accède au fichier son, à une fiche manuscrite rédigée par Jean Dumas, voire à une fiche tapuscrite de la chanson supplémentaire. J'ai ensuite adapté ma méthodologie générale à ce corpus de chansons de tradition orale, avec quelques spécificités. Je vais désormais présenter la façon de procéder que l'expérience a fini par fixer.

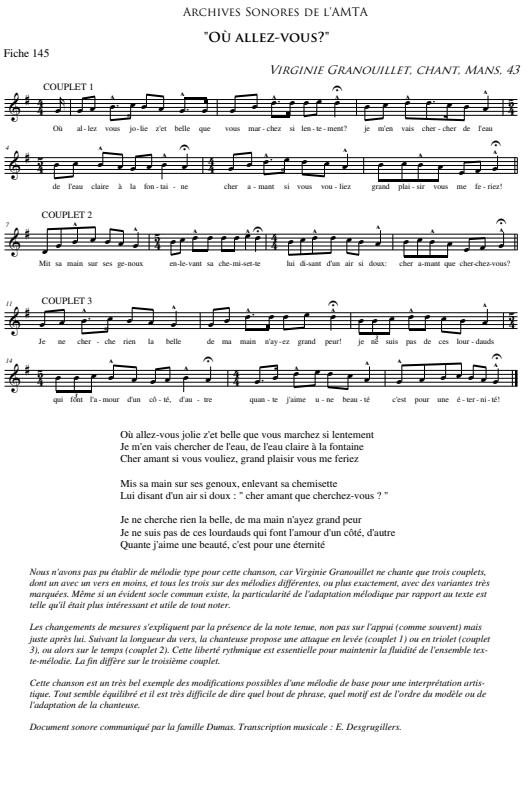
2.3.1 Téléchargement du fichier son et mise en place d'un plan de nommage

Tout d'abord, je téléchargeais le fichier son de chaque chanson depuis la BIPO avant de le renommer. La mise en place d'un plan de nommage s'est révélée complexe. Le fait qu'un plan de nommage régulier n'ait pas été mis en place dès la mise en ligne du FJD ne m'a pas facilité la tâche. J'ai tout d'abord utilisé ce qui semblait correspondre au numéro d'enquête, séparé du second numéro, le numéro de la chanson dans l'enquête, par un tiret bas _. Parfois, le numéro d'enquête est indiqué en cote, mais les cotes ont été renseignées de façon très variable, quand elles l'ont été. Ce numéro d'enquête + numéro de chanson dans l'enquête peut parfois être l'appellation qui s'affiche pour le document sonore sur la page de la chanson en question (notamment dans le nom du fichier audio), il peut aussi être référencé sur la page de l'enquête ou sur la page du

de les convertir sous forme compressée, comme aux formats MP3 ou OGG pour la diffusion sur le web, selon les préconisations de la Très grande infrastructure de recherche (TGIR) Huma-Num (Marcadé et al., 2014 : 74).



(a) Version manuscrite de Jean Dumas



(b) Version tapuscrite

FIGURE 2.4 : *Où allez-vous jolie-z-et belle que vous marchez si lentement ?* (JD_a060_39)

corpus. Il faut donc constamment jongler avec différents niveaux d'archivage, ce qui s'avère assez laborieux. J'ai par la suite décidé d'ajouter JD_a devant tous les fichiers (avec soustraction du premier 0 du numéro d'enquête), comme cela avait été fait pour certaines chansons, peut-être les dernières à avoir été mises en ligne, ce qui s'inscrit dans la logique des corpus annoncés (dont il est fait mention en 3.2.2). Observons quelques exemples :

- la chanson *oEuDOI0887yDqCPm4ku_YGvNn-o92QhH* au téléchargement, nommée *0226* sur le site (il s'agit du numéro de la fiche rédigée par Jean Dumas précédé d'un 0), est renommée en *JD_a063_01* (*0063_01* est indiquée pour la chanson en question sur la page générale du corpus ; voir BIPO, 2022g)
- la chanson *oEuDOI0887wLiW1an7hEIy_vcrDidnLY* au téléchargement, nommée *enq0156 (9)* sur le site, est renommée en *JD_a156_09*.

En outre, le FJD accessible en ligne présente deux défauts majeurs : aucun fichier

descriptif recensant l'ensemble des chansons n'a été constitué et la hiérarchisation des chansons laisse à désirer : le fonds se divise en cinq sous-fonds, dont l'un deux comporte cinq corpus sonores (j'ai transcrit et annoté une grande partie des chansons de quatre d'entre eux) qui se divisent en enquêtes ; les chansons sont également reliées à leur interprète et à un corpus mais certaines disposent d'une page qui leur est propre sur la BIPO, quand d'autres ne sont référencées que sur la page de l'enquête dont elles sont issues. Tout ceci fait que le fonds doit être minutieusement décortiqué pour éviter de passer à côté de certaines productions. Certaines chansons ont pu échapper à mon investigation.

Retenons donc que le traitement du FJD dans le cadre de cette thèse débute avec le téléchargement du fichier son et son renommage. Des inconsistances dans le plan de nommage originel du FJD sur la BIPO ont complexifié mon plan de nommage du fonds. J'ai décidé d'opter pour le modèle suivant : JD_a(numéro de l'enquête, premier zéro soustrait)_(numéro de la chanson dans l'enquête), en espérant que cela soit ce vers quoi tend actuellement le plan de nommage du FJD mis en ligne sur la BIPO. Enfin, l'absence de recensement de toutes les chansons et la hiérarchisation inégale de celles-ci (avec ou sans fiche individuelle) ne facilitent pas cette étape. Pousuivons notre méthodologie de traitement du FJD : nous allons à présent lier un fichier .cha au fichier son téléchargé.

2.3.2 Préparation d'un fichier .cha

Dans cette deuxième étape, un fichier .cha est créé pour travailler sur CLAN, portant le même nom que la chanson du FJD téléchargée. Comme vu en 2.1.3.2, l'objectif principal de cette étape est de renseigner les métadonnées associées à la chanson. Au fil du temps, je me suis créé un fichier .cha modèle avec une en-tête précomplétée. Je travaillais chanteur par chanteur et pour la première chanson d'un interprète, je modifiais mon fichier modèle. Avait été mentionné en 2.1.3.2 l'ajout des métadonnées suivantes : âge[29], sexe présumé et rôle du chanteur, langue de la chanson et nom de la transcriptrice. Les deux premières informations provenaient des fiches de Jean Dumas (voir figure 2.5). Pour le FJD, les autres rôles que chanteur pouvaient être : Interviewer pour l'enquêteur (il s'agit normalement de Jean Dumas) ou Audience pour le public, ce dernier pouvant donc correspondre à des locuteurs distincts pour une même chanson.

[29]Nous avions vu que j'indiquais l'âge des chanteurs de mes corpus en années. De toute façon, Jean Dumas renseigne uniquement l'âge des locuteurs, excepté pour Virginie Granouillet pour laquelle on dispose de sa date de naissance.

J'indiquais également les informations suivantes, issues des mêmes fiches : le prénom et le nom, le métier et le lieu de vie du chanteur (toujours indiqué sous la forme d'un code postal[30]). Je n'indiquais pas les informations liées à notre collecteur comme son âge ou son lieu de vie, même si celles-ci m'étaient connues, car je m'intéressais aux productions linguistiques des chanteurs.

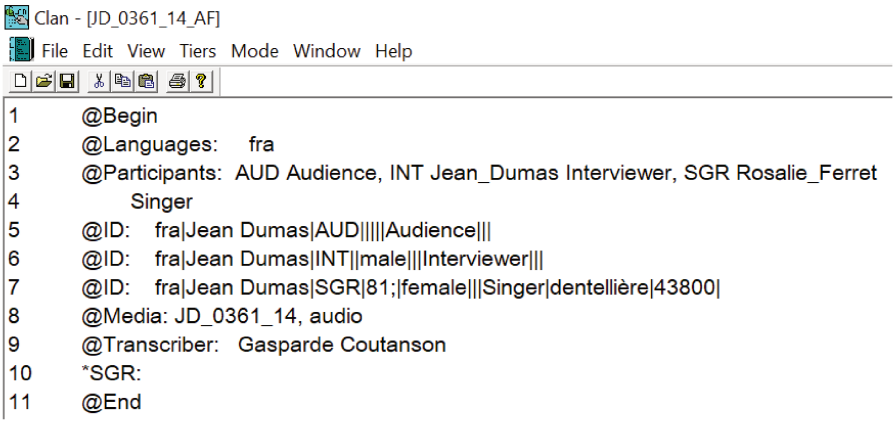


FIGURE 2.5 : En-tête du fichier .cha de la chanson *À Vorey le petit bourg* (JD_a361_14)

Au cours de cette étape a été mentionnée la création d'un fichier .cha portant le même nom que l'audio téléchargé depuis le FJD. Pour résumer, celui-ci doit comprendre les métadonnées suivantes pour ce corpus : prénom, nom, âge, sexe présumé, métier, lieu de vie et rôle de l'interactant ; langue, nom du transcripteur et enfin fichier son à associer. L'étape suivante est celle de l'alignement son/texte.

[30] J'ai décidé d'opter pour l'utilisation de codes postaux car certains logiciels permettent une cartographie automatique des lieux à partir de codes postaux. Ce choix peut se révéler problématique, car on sait que de petites communes proches peuvent avoir le même code postal. C'est par exemple le cas pour le FJD des communes de Rosières et de Chamalières-sur-Loire, qui partagent un même code postal, 43800. Lors d'une analyse ou d'une cartographie des données, deux communes différentes pourraient alors être analysées conjointement au lieu de l'être séparément. Par ailleurs, il faut noter que les communes ont pu changer de nom depuis la collecte, notamment pour éviter les doublons au niveau national (les nouveaux noms sont souvent des versions plus longues). Il peut aussi y avoir eu des réorganisations au sein des communes, par exemple des regroupements. J'ai tenté de retrouver au mieux les lieux de collecte d'origine, mais mes en-têtes pourraient néanmoins présenter des erreurs. Les logiciels de cartographie fonctionnent aussi en partant de noms de lieux (actualisés), sous réserve qu'ils n'évoluent plus (aucune garantie ici), mais il me semblait que l'utilisation de codes postaux réduirait au mieux la marge d'erreur possible.

2.3.3 Alignement son/texte

Pour cette troisième étape, il faut repartir du fichier .cha créé dans l'étape précédente, avec métadonnées renseignées et audio associé. Il faut ensuite baliser le document selon la méthode indiquée en 2.1.3.3.

Pour le FJD, j'ai choisi de le faire en m'appuyant sur les pauses saillantes à l'écoute, notamment dues aux inspirations des chanteurs, ainsi qu'à chaque tour de parole. J'ai décidé d'opter pour ce type de découpage en total désaccord avec le découpage habituellement utilisé pour transcrire des chansons, soit le découpage en vers, et ce choix est lourd de conséquences. Tout particulièrement, les corpus ainsi transcrits seront difficilement réutilisables en ethnomusicologie ou en linguistique pour travailler par exemple sur des questions de métrique. Il sera également impossible de vérifier certaines affirmations telles que celles que nous avions mentionnées en 1.2.3.4.2 : « la liaison ne se fait pas entre deux vers ». Pour autant, cette segmentation me semblait plus simple (pour éviter, par exemple, de confondre une césure avec une fin de vers) et permettait de rendre au mieux des productions spontanées présentant par exemple des corrections auto-initiées. J'ai toléré une seule exception à ce découpage : j'ai généralement regroupé les liaisons après pauses ou inspirations avec la tire précédente, afin que lors de l'extraction des données, le M1 soit automatiquement proposé. Les exemples 76 et 77 tiennent lieu d'illustration : l'annotation en 77 est préférée à celle en 76, malgré les inspirations entre les *vous* et les *aimiez* (première pause de 0,36s et deuxième de 0,39s[31]). Un découpage à la pause, reposant notamment sur les inspirations des chanteurs, entraîne parfois des transcriptions orthographiques surprenantes, notamment quand une inspiration se produit à l'intérieur d'un mot (exemple 78).

(76) *SGR : je l'aimerai si tendrement comme vous.
*SGR : [z] aimiez mon père comme vous.
*SGR : [z] aimiez mon père.
(JD_a205_02, 43810, Marie Soulier)

(77) *SGR : je l'aimerai si tendrement comme vous aimiez mon père comme vous aimiez mon père.
(JD_a205_02, 43810, Marie Soulier)

(78) *SGR : en entrant la porte se ferme il ou(blie).
*SGR : (ou)bli(e) la clé par dehors.
(JD_a205_07, 43810, Marie Soulier)

[31] durées obtenues à l'aide du logiciel Praat

On réalise donc un alignement son/texte pour cette troisième étape. Pour le FJD, le découpage en tires se fonde sur les pauses saillantes à l'écoute et les tours de paroles, plutôt que sur les vers, ce qui pourrait complexifier de potentielles réutilisations du corpus. Dans l'étape suivante, nous présenterons les spécificités de transcription et d'annotation du FJD, en mentionnant au passage de multiples spécificités phoniques de ce corpus.

2.3.4 Transcription et annotation du document sonore

Lors de cette quatrième étape, l'audio est transcrit et annoté sur le document .cha, uniquement pour les parties chantées (et non les passages parlés[32]). La transcription est faite sans consulter les fiches fournies sur la BIPO. Dans cette partie, je vais commencer par présenter deux annotations spécifiques au FJD.

Je vais revenir dans un premier temps sur l'annotation des schwas dans le FJD. Pour les chansons de tradition orale, j'ai placé tous les ‹e› graphiques n'appartenant pas à des multigrammes (du type ‹eu›, ‹ei›...) entre parenthèses s'ils n'étaient pas réalisés, ce qui signifie que tout ‹e› graphique sans parenthèses correspond à un schwa prononcé. Les schwas non étymologiques en frontière de mots ont pour leur part été annotés [*@], le symbole @ correspondant au schwa en caractères SAMPA, et les schwas internes au mot n'ont pas été annotés sur la tire principale mais sur la tire dépendante de transcription phonétique, où la tire principale a entièrement été transcrite en API.

Dans un second temps, arrêtons-nous plus longuement sur la notation des (non)-réalisations de consonnes dans le FJD. Quatre cas sont à distinguer :

- la non-réalisation de certaines consonnes, bien souvent prononcées aujourd'hui sous l'influence de graphies étymologisantes, tardivement réintroduites

  On trouve par exemple les archaïsmes suivants :

(79) *SGR : on parle dans la ville d'un ca(p)tif [*@] si beau.
(JD_a063_20, 43810, Virginie Granouillet)

(80) *SGR : bien vit(e) et prom(p)tement.
(JD_a165_07, 43800, Germaine Varenne)

[32] De cette façon, une comparaison entre voix chantée et voix parlée n'est pas possible pour le FJD, bien que l'alignement son/texte ait bien été réalisé pour chaque intervention. Par sécurité, les vers (comme pour JD_a330_16, 28-48) et les chansons (ex : JD_a347_08) déclamés ont été exclus de l'analyse.

- de la variation dans la réalisation de groupes consonantiques

  Notamment, une prononciation régionale et/ou populaire revient fréquemment dans le FJD, celle de réaliser les [lj] [j], comme pour bate(l)ière (JD_a165_38, 44), cava(l)iers (JD_a062_07, 33), esca(l)iers (JD_a359_01, 29), geô(l)ier (JD_a165_12, 25), (l)iards (JD_a156_05, 33), mi(l)ieu (comme en 81), mi(ll)ions (JD_a075_05, 40), mi(ll)ionnaire (JD_a060_32, 43), mobi(l)ier (JD_a060_08, 31-32), particu(l)ière (JD_a075_22, 81) ou encore sou(l)iers (JD_a061_12, 27).

(81) *SGR : ne fur(e)nt pas dans la barque ne furent pas au mi(l)ieu de l'eau.
(JD_a060_39, 43810, Virginie Granouillet)

Dans l'exemple suivant, c'est cette fois-ci la séquence [tr] qui est réduite à [t] :

(82) *SGR : quan(d) l(e) prê t(r e) fut arrivé.
(JD_a061_11, 43810, Virginie Granouillet)

- de la variation dans la réalisation des consonnes finales de certains mots, quel que soit le contexte de production du mot (à la pause, devant initiale vocalique ou devant initiale consonantique)

  On observe alors deux formes possibles pour un même mot : prononcé sans consonne finale et avec consonne finale. On parle de *doublet*. Dans le premier cas, les consonnes finales ne sont pas lexicalisées alors que dans le second cas, elles le sont. Néanmoins, il est nécessaire de garder à l'esprit qu'il pourrait exister de la variation chez un même locuteur dans la réalisation de cette consonne finale et ce toujours sans que cela soit lié à un contexte de production en particulier. Les mots impliqués ne sont pas forcément ceux habituellement mentionnés dans la littérature. Walter (1988) cite à titre d'illustration les termes suivants : *almanach*, *ananas*, *août*, *but*, *cerf*, *chenil*, *circonspect*, *exact*, *fait (un)*, *gril*, *nombril*, *persil*, *sourcil* et *suspect*. On retrouve par exemple une non-réalisation de la consonne finale pour le mot *août* dans le FJD :

(83) *SGR : aoû(t) nous produit.
(JD_a168_04, 63550, Marie Rigodias)

Ici, *août* est prononcé [au]. Cette non-réalisation du [t] final est considérée comme un archaïsme ou un régionalisme. On peut également observer d'autres non-réalisations dans le FJD telles que pour *lys*, prononcé [li], *ours*, prononcé [ur], *luth*, prononcé [ly], *blanc-bec*, prononcé [blɑ̃be] (ces quatre cas en fin de vers peuvent être vérifiés à la rime) ou *mérinos*, prononcé [meʁino] :

(84) *SGR : au jardin de son père par dessous ces fleurs de ly(s).
(JD_a060_10, 43810, Virginie Granouillet)

(85) *SGR : des lions et des our(s).
(JD_a062_01, 43810, Virginie Granouillet)

(86) *SGR : un drôl(e) de joueur de lu(th) un drôl(e) de joueur de lu(th).
(JD_a328_05, 43800, André-Joseph Aguilhon)

(87) *SGR : si je l'ai fait je l'ai bien fait je n'en suis pas de ces blancs + be(c)s.
(JD_a165_01, 43800, Jean Déreil)[33]

(88) *SGR : un(e) en mérino(s) bleu et l'autr(e) en satin rose chan(te).
(JD_a066_25, 43800, Rosalie Ferret)

Au contraire, on trouve des réalisations archaïques des consonnes finales de mots tels que *ceux*, prononcé [søs], *gens*, prononcé [ʒɑ̃s] ou *tabac*, prononcé [tabak] :

(89) *SGR : d(e) nos chèr(e)s maîtresses ceux qui en auront.
(JD_a168_03, 63550, Marie Rigodias)

(90) *SGR : c'était pour saluer les gens du mariage.
(JD_a060_27, 43810, Virginie Granouillet)

(91) *SGR : tu nous vendras du la bière tu nous vendras du tabac [*@] de l'eau + d(e) + vi(e) aux soldats.
(JD_a165_08, 43800, Germaine Varenne)

Certaines réalisations apparaissent dans un contexte très spécifique, comme en 92 et 93, où des [s*] sont produits à la finale de verbes conjugués à la 2.SG du futur (simple ou antérieur), devant l'enclitique *tu*, postposé lors d'une inversion sujet/verbe.

(92) *SGR : camarade m'en donneras [s*] tu tout à l'entour du cul ?
(JD_a064_18, 43800, Rosalie Ferret)

(93) *SGR : oh@i malheureux qu'auras [s*] tu fait tu auras déshonoré ma fille.
(JD_a160_05, 43800, Jean-Pierre Arnaud)

Si je savais que ces mots variaient ou si je l'avais repéré dans le FJD, j'annotais alors systématiquement une réalisation sans parenthèses et une non-réalisation avec parenthèses.

[33] Voir aussi la version de Jean-Pierre Arnaud en JD_a160_05

- de la variation dans la réalisation des consonnes finales de certains mots, en fonction du contexte de production du mot

  Prenons l'exemple de *quand/quant* : devant initiale consonantique, la consonne finale n'est habituellement pas réalisée :

(94) *SGR : de grand matin quan(d) petit Pierr(e) se lève.
(JD_a347_05, 43200, Séraphine Chaize)

J'ai fait le choix de noter systématiquement entre parenthèses ces cas-là pour indiquer que dans mes corpus, ces mots ne sont pas de simples mots pouvant être liaisonnés et qu'on trouvera des réalisations ailleurs que dans un contexte de liaison, c'est-à-dire devant initiale vocalique, où une liaison peut être produite :

(95) *SGR : quand il vient sur les huitièmes je me rends au pays.
(JD_a060_31, 43810, Virginie Granouillet)

Elle est indiquée lors de la phase d'annotation de notre corpus. Parfois, la consonne finale peut en effet être prononcée même devant initiale consonantique :

(96) *SGR : quand [t*] c'était pour la vieille.
*SGR : le moulin n'allait pas du@si clin@si du@si maniclin@si.
(JD_a350_01, 43200, Séraphine Chaize)

Le mot est alors transcrit sans parenthèses et pour marquer l'atypicité de la réalisation, il est annoté avec le postcode [x*] avec x = consonne finale réalisée. Ce code ne sert pas à noter les consonnes finales habituelles des mots *plus*, *tous* ou encore *six* ou *dix*. Il a l'inconvénient d'être aussi utilisé pour les cas incertains de liaisons. Voyons à présent quelques cas plus complexes encore.

Le premier cas que nous aborderons sera celui du mot *plus*. Il est bien connu qu'il peut avoir aussi bien une polarité négative que positive. Dans le premier cas, il sera alors habituellement prononcé *plu(s)* :

(97) *SGR : de crédit nous en faisons plu(s) le curé nous l'a défendu.
(JD_a361_14, 43800, Rosalie Ferret)

Une liaison peut tout de même être réalisée (ce qui est le cas pour les deux polarités) :

(98) *SGR : madam(e) la marié(e) vous n'irez plus [z] au bal.
(JD_a159_07, 43800, André-Joseph Aguilhon) : polarité négative

(99) *SGR : les quatre les plus [z] ivrognes porteront le drap de mort.
(JD_a330_04, 43800, Henriette Usson) : polarité positive

Pour la polarité positive (hors cas de liaison déjà cité), la réalité est plus complexe, la consonne finale pouvant être ou non réalisée. Dans notre corpus, on trouvera principalement des non-réalisations :

(100) *SGR : le plu(s) petit s'est avancé.
(JD_a060_34, 43810, Virginie Granouillet)

On peut toutefois noter certaines réalisations :

(101) *SGR : bien le bonjour mère l'abbesse laisseriez vous n'en plus parler ?
(JD_a165_18, 43800, Germaine Varenne)

En français contemporain, toujours en ce qui concerne la polarité positive, certaines constructions seront majoritairement produites en prononçant la consonne finale de *plus* comme dans les constructions *plus de* + X, *plus que* + X ou *plus* + X + *que* + X ou encore lorsqu'il s'agit des locutions adverbiales *de plus* ou *en plus* (ces dernières sont absentes du FJD).

Le second cas que nous aborderons est aussi bien connu et touche aux nombres cardinaux en français contemporain. On distinguera quatre types de fonctionnement. L'ensemble sera résumé dans le tableau 2.2 un peu plus bas. Pour les numéraux composés, il importe de prendre en compte le dernier nombre.
Pour commencer, certains nombres cardinaux, appartenant au premier type, ont une forme fixe avec consonne finale, voire deux formes en cas de pataquès, que ce soit devant initiale consonantique, à la pause ou devant initiale vocalique. C'est le cas de *quatre* (avec simplification du groupe consonantique possible), *sept*, *neuf* (sauf devant initiale vocalique quand le M2 est fréquent comme avec *heures* ([nœvœʁ]) et *ans* ([nœvɑ̃]) où la consonne voisée [v] est produite lors de l'enchaînement plutôt que son équivalente non voisée [f]), *onze*, *douze*, *treize*, *quatorze*, *quinze*, *seize*, *trente*, *quarante*, *cinquante*, *soixante*[34] et *mille*. De la variation peut apparaître si un pataquès est produit : en contexte de liaison, un [z] peut être inséré devant l'initiale vocalique, après le schwa (exemple 102) ou après simplification du groupe consonantique et non-réalisation du schwa (exemple 103) pour *quatre* ou après la finale consonantique pour *mille* (exemple 104). Notons tout de même que de l'ancien français au XVI^{ème} s., « *mille* pouvait être mis au pluriel quand

[34] C'est également le cas pour les variantes *septante*, *huitante* et *nonante* de Belgique et de Suisse romande.

il était multiplié », même si cet emploi était rare (Marchello-Nizia et al., 2020 : 730). Par la suite, sous la pression de grammairiens, *mille* est devenu invariable (Marchello-Nizia et al., 2020 : 730).

(102) *SGR : par quatre [*z] officiers marine.
(JD_a062_03, 43810, Virginie Granouillet)

(103) [...] et y avait euh y avait David, mon pote, et y avait genre deux trois quat(re) [*z] autres gens [...].
(ce, 2018, Bilal Hassani)

(104) C'est aussi une place pour chacun et je compte sur vous pour euh donner leur place dans l'école aux quatre cent mille [*z] enfants et adolescents qui sont en situation de handicap [...].
(ce, 2020, Emmanuel Macron)

En cas de pataquès, pour les nombres dont la consonne finale est [z], une géminée est alors produite (Gadet, 1989 : 56), comme dans l'occurrence *douze enfants* [duz.zɑ̃.fɑ̃] relevée en français louisianais de Golden Meadow par Blainey (2013 : 208). Dans le FJD, on trouve plus de variation car *neuf* peut être produit sans consonne finale devant initiale consonantique, ce qui constitue un archaïsme. C'est systématique pour les occurrences de *neuf mois*, prononcées [nømwa] (cette collection d'exemples est exhaustive) :

(105) *SGR : au bout de neu(f) mois de mariag(e) m(e) voilà [*z] un enfant.
(JD_a062_06, 43810, Virginie Granouillet)

(106) *SGR : au bout de neu(f) mois tout au plu(s).
(JD_a062_08, 43810, Virginie Granouillet)

(107) *SGR : au bout de neu(f) mois tout au plu(s) mon ami au bout de neu(f) mois tout au plu(s).
(JD_a063_26, 43810, Virginie Granouillet)

(108) *SGR : au bout d(e) huit à neu(f) mois tout au plu(s).
(JD_a168_02, 63550, Marie Rigodias)

Pour *sept*, on sait que le [t] final n'a pas toujours été prononcé devant initiale consonantique. Cependant, l'homophonie avec *ces* rend complexe l'étape de transcription, comme en 109 :

(109) *SGR : sept ans je l'ai cherchée en sep(t) différents lieux.
(JD_a063_20, 43810, Virginie Granouillet)

J'ai ici opté pour la transcription *sept* plutôt que *ces* car les lieux n'avaient pas été mentionnés auparavant.

Les nombres cardinaux appartenant au second type présentent quant à eux une, deux, voire trois formes : il s'agit de *cinq* et *huit*. Dans le premier cas, la consonne finale du nombre est prononcée quel que soit le contexte, alors que dans le second cas il existe une forme sans consonne finale devant initiale consonantique et une forme avec consonne finale à la pause et devant initiale vocalique. De la même façon que pour *neuf*, la fréquence du couple M1 + M2 pourrait expliquer la réalisation ou non-réalisation de la finale [k] de *cinq* devant initiale consonantique (par exemple, *cinq minutes* est généralement réalisé [sɛ̃minyt]). Dans le FJD, *cinq* précédant un nom à initiale consonantique est toujours réalisé sans consonne finale : *cin(q) sous*, *cin(q) prêtr(e)s*, *cin(q) fois*, *cin(q) coups* et *cin(q) tours* (cette collection d'exemples est exhaustive). Les pataquès touchant ces numéraux pourraient être insérés entre la consonne finale et l'initiale vocalique ou substitués à la consonne finale (notamment pour *huit*), toujours devant initiale vocalique, comme dans l'exemple 110, issu de PFC :

(110) [...] ben elle avait eu hui(t) [*z] enfants, pis elle est morte avec le dernier bébé, elle était pas supposée d'en avoir d'autres [...].
(PFC, Peace River (Canada, Amérique), 2006, caalm1)

Le troisième type concerne les nombres cardinaux qui alternent entre deux formes, une devant initiale consonantique et à la pause, sans consonne finale, et une devant initiale vocalique, avec liaison, voire trois en cas de pataquès. C'est le cas de *un*, *deux*, *trois*, *vingt* et *cent*. Notons que *vingt* a la particularité d'être prononcé avec consonne finale dans les composés numéraux *vingt* + X tels que *vingt et un* ou *vingt-deux* mais pas dans les composés numéraux *quatre-vingt* + X. Historiquement, le dernier chiffre était souvent coordonné : *dis et set*, *vingt et quatre*..., avant que la juxtaposition ne s'impose au XVII^{ème} s., à l'exception des nombres dont le dernier chiffre est *un* ainsi que pour *soixante et onze* (cf Marchello-Nizia et al., 2020 : 729). La consonne finale de *vingt* est parfois lexicalisée dans certains régiolectes. En cas de pataquès, un [z] est substitué à la consonne de liaison, comme pour *vingt* (exemple 111) ou *cent* (exemple 112).

(111) MA1 : Ah les grandes familles là !
GT1 : Des vingt [*z] enfants !
(PFC, Saguenay (Canada, Amérique), 2006, cqbgt1)

(112) [...] Pacs a dû mettre cen(t) [*z] euros de sa poche là.
(PFC, Liège (Belgique, Europe), 2006, blajv1)

Enfin, d'autres nombres appartenant à un quatrième type ont encore un fonctionnement plus complexe et alternent entre trois formes, une devant initiale consonantique, sans consonne finale, une à la pause, avec consonne finale et une devant initiale vocalique, avec consonne de liaison phonétiquement différente de la consonne produite à la pause, voire quatre formes en cas de pataquès. C'est le cas de *six* et *dix*. Devant initiale consonantique, il est tout de même possible de trouver des réalisations de consonnes finales semblables aux formes à la pause :

(113) *SGR : après dix [s*] campagnes passées.
(JD_a060_18, 43810, Virginie Granouillet)

Pour ce quatrième type, on pourrait aussi trouver des enchaînements en [s] devant initiale vocalique au lieu de liaisons en [z], comme le mentionne Delattre (1956 : 53) pour le cas particulier des dates (nombre cardinal + nom de mois) : [ləsisavʁil] ou [ləsizavʁil] pour *le six avril* ?
Comme annoncé plus tôt, le tableau 2.2 récapitule les différentes prononciations que peuvent prendre les nombres en français contemporain. Ce tableau reprend des données introspectives ainsi que, pour les pataquès y étant indiqués, les données issues de collections d'exemples et de corpus susmentionnées.

Le troisième cas que nous aborderons est celui de *tous*, qui peut être soit un adjectif[35], soit un pronom en français. Actuellement, lorsque *tous* est adjectif, il n'est pas prononcé comme dans :

(114) *SGR : oh@i oui oh@i oui bon capitaine tou(s) mes rasoirs sont prêts pour vous.
(JD_a330_02, 43800, Henriette Usson)

Dans notre corpus de chansons, on trouve pourtant des archaïsmes tels qu'en 115 et 116 :

(115) *SGR : sont troi(s) beaux capitaines.
tous [s*] troi(s) voudraient lui fair(e) la cour.
(JD_a330_02, 43800, Henriette Usson)

(116) *SGR : quand ils l'eurent bien tué(e) tous [s*] troi(s) voulurent bien souper.
(JD_a361_21, 43800, Rosalie Ferret)

[35] Certaines grammaires le considèrent comme un déterminant.

Type	Nombre	Forme devant initiale consonantique	Forme à la pause	Forme devant initiale vocalique	Exceptions
Type 1	(+)quatre	[katʁə] ~ [katʁ] ~ [kat]	[katʁ] ~ [kat]	[ka.tʁ] ~ [kat.z]	/
	(+)sept	[sɛt]	[sɛt]	[sɛ.t]	/
	(+)neuf	[nœf]	[nœf]	[nœ.f]	[nœ.v] devant *heures*, *ans* (M2 fréquent)
	(+)onze	[ɔ̃z]	[ɔ̃z]	[ɔ̃.z]	
	(+)douze…	[duz]	[duz]	[du.z]	/
	(+)trente	[tʁɑ̃t]	[tʁɑ̃t]	[tʁɑ̃.t]	
	(+)qua-rante…	[kaʁɑ̃t]	[kaʁɑ̃t]	[kaʁɑ̃.t]	
	(+)mille	[mil]	[mil]	[mi.l] ~ [mil.z]	/
Type 2	(+)cinq	[sɛ̃] ~ [sɛ̃k]	[sɛ̃k]	[sɛ̃k]	/
	(+)huit	[ɥi] ~ [ɥit]	[ɥit]	[ɥit] ~ [ɥi.z]	/
Type 3	(+)un	[œ̃] ~ [ɛ̃]	[œ̃] ~ [ɛ̃]	[œ̃.n] ~ [ɛ̃.n]	/
	(+)deux	[dø]	[dø]	[dø.z]	/
	(+)trois	[tʁwa]	[tʁwa]	[tʁwa.z]	/
	(+)vingt	[vɛ̃]	[vɛ̃]	[vɛ̃.t] ~ [vɛ̃.z]	nombres de 21 à 29 : [vɛ̃teɛ̃] ~ [vɛ̃teœ̃], [vɛ̃tdø], [vɛ̃ttʁwa], vɛ̃tkat(ʁ)(ə)], [vɛ̃tsɛ̃k], [vɛ̃tsis], [vɛ̃tsɛt], [vɛ̃tɥit], [vɛ̃tnœf]. (+)vingts (après un multiplicateur, en fin de nombre) devant initiale vocalique (liaison) : [vɛ̃.z]
	(+)cent	[sɑ̃]	[sɑ̃]	[sɑ̃.t] ~ [sɑ̃.z]	(+)cents (après un multiplicateur, en fin de nombre) devant initiale vocalique (liaison) : [sɑ̃.z]
Type 4	(+)six	[si] ~ [sis]	[sis]	[si.z]	/
	(+)dix	[di] ~ [dis]	[dis]	[di.z]	/

TABLEAU 2.2 : Variation dans la prononciation des nombres en français contemporain selon leur contexte de réalisation

Notons le postcode [s*] qui souligne le caractère inhabituel de ces réalisations. Dans le FJD, on ne trouve de semblables exemples que pour *tous trois* et jamais pour *tous deux*. Ici, l'adjectif *tous* est prononcé de la même façon que le pronom *tous*, avec consonne finale (deuxième *tous* de l'exemple 117) :

(117) *SGR : mais tou(s) tes camarades se sont tous mariés.
(JD_a060_04, 43810, Virginie Granouillet)

Ici, pas de postcode car ces réalisations sont typiques. Notre corpus nous fournit à contrario des exemples de non-réalisation de la consonne finale du pronom *tous*, ce qui est inhabituel pour un auditeur habitué au français contemporain (deuxièmes *tous* des exemples 118 et 119) :

(118) *SGR : la guerr(e) est déclarée pa(rtout) a été calmé(e) partout.
*SGR : tou(s) nos soldats resteront tou(s) revenez !
*SGR : filles de Rennes !

(JD_a060_32, 43810, Virginie Granouillet)

(119) *SGR : tou(s) les voisins courriront tou(s).
(JD_a155_03, 43810, Maria Drevet)

(120) *SGR : tou(s) i(l)s sont à vot(re) service jusqu'au plu(s) beau.
(JD_a063_32, 43810, Virginie Granouillet)

Enfin, on trouve des cas d'enchaînements avec voisement de la consonne enchaînée de *tous*, comme dans les exemples 121 et 122 :

(121) *SGR : oh@i i(l)s les ont mis tous en poussières.
(JD_a062_03, 43810, Virginie Granouillet)

(122) *SGR : ils s'en vont tous en cherchant le divertissement.
(JD_a329_02, 43350, Joseph Bonnefoux)

Enfin, je terminerai en présentant une difficulté de transcription concernant *tous* que j'ai rencontrée dans certains vers. En effet, *tous* adjectif est homophone de *tout* adjectif (ou déterminant). Cela peut poser problème devant initiale consonantique. À titre d'illustration, l'exemple 123 aurait pu être transcrit entièrement au singulier (cet exemple présenterait alors un pataquès) :

(123) *SGR : tou(s) garçons [z] à marier cherchent fortune.
(JD_a063_27, 43810, Virginie Granouillet)

Mais des exemples comme en 124, avec un verbe marquant oralement le pluriel (*tâchez*), m'ont encouragée à transcrire *tous* au pluriel. En outre, produire du silence plutôt que du bruit est préférable pour des données plus fiables.

(124) *SGR : oh@i tou(s) garçons à marier.
*SGR : tâchez d'en trouver la pareille !
(JD_a060_07, 43810, Virginie Granouillet)

Lors de l'annotation des liaisons et pataquès dans le FJD, contrairement à celle du corpus des tops 1, j'ai adopté l'annotation miminale qui consiste à n'indiquer que les sandhis inattendus hors contexte de liaison, qui ne pourront autrement pas être récupérés par la suite. La lecture des transcriptions s'en trouve encore plus allégée.

Pour le FJD, en cas de doublon, j'ai transcris les deux versions car deux performances uniques ne peuvent être en tous points similaires. Cela pose problème lors de l'analyse, nous reviendrons sur ce point par la suite. Si le plan de nommage des déposants ne reliait pas les documents en les nommant différemment, ces doublons étaient

plus difficilement repérables. Pour les doublons présentés sous une même appellation, j'ai rajouté un numéro à la fin du nom du fichier (_1, _2, _3...).

Les fiches manuscrites de Jean Dumas auraient pu être océrisées, mais je n'ai pas opté pour cette solution, ne maîtrisant pas les techniques d'océrisation, sachant également qu'un tel logiciel devait être entraîné sur l'écriture de Jean Dumas, que l'organisation-même des fiches (titre suivi d'une partition puis organisation en paragraphes amassés sur l'axe horizontal et vertical ; cf fiche manuscrite à gauche de la figure 2.4) poserait problème lors de la transposition sur les fichiers .cha, organisés différemment (tires amassées sur l'axe vertical), que l'objectif du folkloriste ou de l'ethnomusicologue peut différer de celui du linguiste, en cela qu'à la manière des interactionnistes, la moindre petite disfluence nous importe (cf exemple de réparation dans l'exemple 118, p.172) et enfin qu'il aurait fallu reprendre les transcriptions de Jean Dumas océrisées de façon poussée afin de les adapter aux conventions de transcription de CLAN. J'aurais encore pu réutiliser au moins les .pdf tapuscrits constitués pour le corpus portant sur le répertoire de Virginie Granouillet, ce que je n'ai pas fait, même si un essai aurait pu se révéler prometteur.

Ainsi, lors de cette étape, l'audio de chaque performance est transcrit et annoté en respectant les conventions générales du logiciel CLAN vues en 2.1.3.4, sans l'aide de la ou des fiches liées à chaque chanson sur la BIPO, et sans qu'elles ne soient non plus océrisées afin de raccourcir le temps de transcription. J'ai fait part dans cette partie d'annotations que j'avais retenues pour le FJD, concernant les schwas et la (non-)réalisation de certaines consonnes. J'ai pris le temps de détailler quatre cas de variation pour ces dernières, en insistant tout particulièrement sur les consonnes finales, avec une variation non liée à un contexte de production spécifique (j'en ai alors profité pour présenter quelques archaïsmes qu'on peut noter dans le FJD) et une variation liée à un contexte de production spécifique (ont été mentionnés les cas de *quand/quant*, de *plus*, des nombres cardinaux et de *tous*). L'annotation des liaisons et pataquès est minimale pour ce corpus, seuls les sandhis inattendus hors contexte de liaison étant annotés durant cette étape. Pour terminer, une réécoute de l'audio est effectuée en vérifiant ce qui a été produit.

2.3.5 Réécoute et vérification

La dernière étape consistait à vérifier le fichier .cha produit en écoutant une dernière fois l'audio associé. Pour le FJD, j'ai effectué deux réécoutes :

- une première immédiatement après chaque chanson, qui a permis des corrections

au niveau de la transcription, des annotations et du balisage et surtout de comparer ma propre production à celle de Jean Dumas (voir figure 2.6), par exemple en ce qui concerne certains archaïsmes qui m'auraient échappé

- une seconde après la transcription de l'ensemble des chansons, afin de vérifier en outre la bonne normalisation de certaines transcriptions et annotations pour lesquelles j'ai tranché au fil des chansons.

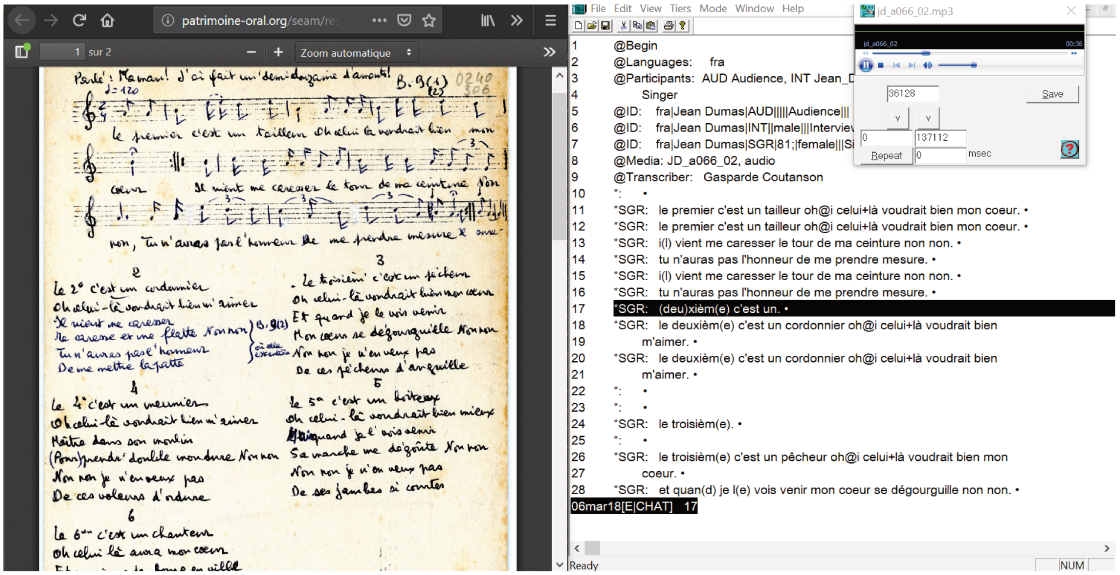


FIGURE 2.6 : Vérification du fichier .cha et comparaison avec la fiche manuscrite de Jean Dumas

Chose qui ne m'est pas venue à l'esprit au commencement de cette entreprise de transcription et d'annotation du FJD, j'ai décidé tardivement de chronométrer le temps qui était nécessaire à cette tâche. Je comptabilisais conjointement les quatre premières étapes, soit l'enregistrement du fichier son avec renommage, la création d'un fichier .cha avec modification des métadonnées au besoin, l'alignement son/texte puis une première transcription/annotation et séparément chaque vérification de la cinquième étape. La durée moyenne des quatre premières étapes de transcription/annotation est de 13min 28s (sur 188 chansons pour lesquelles j'ai pris le temps de me chronométrer) et celle de la cinquième étape est de 05min 17s (sur 188 chansons) pour la première réécoute et de 10min 55s (sur 59 chansons) pour la seconde réécoute, soit environ 30min 48s par chanson (sur 59 chansons).

Dans cette dernière étape, la transcription et l'annotation effectuées dans le fichier .cha sont donc vérifiées en deux fois en les comparant à la fiche manuscrite produite par Jean Dumas. Toutes les étapes de la méthodologie nécessitent en moyenne 30min de travail par chanson.

Récapitulons : quelle méthodologie a été mise en place pour le traitement du fonds Jean Dumas ? Tout d'abord, chaque audio était récupéré sur la BIPO, renommé et un fichier .cha associé était créé. Dans ce fichier, je renseignais les métadonnées avant d'aligner le son et le texte, avec un découpage au niveau des pauses saillantes à l'écoute. Pour finir, je transcrivais et j'annotais les chansons en respectant les conventions de CLAN, avant d'associer chaque tire à un participant. J'ai dû procéder à certains choix de transcription et d'annotation qui ont été présentés. Pour terminer, je vérifiais les fichiers .cha deux fois à l'aide de la fiche manuscrite rédigée par Jean Dumas.

Dans ce chapitre, j'ai dans un premier temps annoncé la méthodologie générale que j'avais mise en place dans le cadre de cette thèse, puis dans un second temps présenté les adaptations qui ont été nécessaires pour le corpus de tops 1 que j'ai compilé, transcrit et annoté ainsi que pour le fonds Jean Dumas que j'ai réutilisé, transcrit et annoté. Le chapitre suivant sera consacré à une présentation de ces deux corpus principaux sur lesquels se fonde cette thèse. En guise d'introduction, les premières archives sonores à avoir été enregistrées pour le français, les Archives de la Parole, seront présentées.

Chapitre 3

Présentation du fonds Jean Dumas et du corpus de tops 1

3.1 Ferdinand Brunot et les Archives de la Parole : premiers enregistrements de la langue française

Après la fondation des Phonogrammarchiv de Vienne en 1899 par Sigmund Exner puis des Phonogramm-Archiv de Berlin en 1905 par Carl Stumpf, c'est au tour de Paris de se doter d'archives sonores institutionnelles avec la fondation des Archives de la parole en 1911 par Ferdinand Brunot (Cordereix, 2001 : 40,42 ; Kursell et al., 2021 : 4). Les Archives de la Parole se subdivisent en plusieurs sections (Cordereix, 2001 : 42-43) :

- I pour interprètes : il s'agit d'enregistrements de prononciation normative du français
- O pour orateurs : il s'agit d'écrits oralisés produits par des personnalités de l'époque
- L pour langues : il s'agit d'enregistrements en langues étrangères à visée didactique
- M traite de pathologies du langage
- D pour dialectes : il s'agit d'enregistrements de dialectes, souvent français, avec pour objectif la constitution d'un atlas sonore.

Contrairement aux autres sections, pour recueillir les données dialectologiques des Archives de la parole, il est nécessaire de faire du terrain. Ce n'est pas la première fois que des appareils enregistreurs sont utilisés lors d'enquêtes de terrain (l'abbé Rousselot en avait utilisés pour sa thèse soutenue en 1892), mais ici, la méthodologie mise en

place est tout à fait novatrice : grâce à Émile Pathé, l'équipe se dote d'un phonographe enregistrant sur disque plat, et par ailleurs d'une automobile (Cordereix, 2001 : 40, 45-46). Trois missions phonographiques ont lieu entre 1912 et 1913 dans les Ardennes franco-belges, dans le Berry puis dans le Limousin (Cordereix, 2001 : 43). Si l'objectif est d'enregistrer certains dialectes menacés, l'entreprise est vue comme un élément de documentation de l'histoire du français et Ferdinand Brunot, pour des raisons politiques (il est farouchement républicain), est un fervent partisan du monolinguisme et conserve une vision archaïsante des dialectes et de leurs locuteurs (Cordereix, 2001 : 39, 49-50). Cette entreprise était novatrice également car les enquêtes phonographiques suivantes, ethnomusicologiques, ne verront le jour que vingt-six ans plus tard, et il faudra attendre quarante ans pour que les enquêtes linguistiques phonographiques deviennent plus courantes (Cordereix, 2001 : 52).

Les premières archives sonores françaises, les Archives de la Parole, ont donc vu le jour à la même époque que d'autres projets similaires sur le continent européen. Les premiers enregistrements qui y figurent concernent de la prononciation normative du français, des écrits oralisés de personnalités de l'époque, des langues étrangères, des pathologies du langage et des dialectes. Pour ces derniers, l'enquête de terrain est nécessaire ; les enquêteurs sont véhiculés et utilisent l'enregistrement sonore. Nous allons à présent longuement présenter un fonds constitué de plusieurs corpus, le fonds Jean Dumas (FJD), que j'ai transcrit et annoté dans le cadre de cette thèse, avec un premier point où j'explique pourquoi c'est ce fonds que j'ai choisi.

3.2 Fonds Jean Dumas

3.2.1 Choix d'un corpus de chansons de tradition orale

Commençons par recenser les différentes appellations qui peuvent être données à notre objet-chanson. J'ai opté pour celle de **chanson de tradition orale** ou « [o]bjet-chanson dont l'identité de l'auteur et du compositeur de musique est inconnue et qui circulait généralement par la transmission orale – d'où les multiples versions des chansons » (Surmont, 2010 : 153). Celle de ***chanson populaire***, ou « chan[t] littérair[e] ou folkloriqu[e] aim[é] du public », interprété par « un artiste qui plaît au peuple »[1] (Laforte, 1976 : 6) et celle de ***chanson traditionnelle*** ou « chanson transmise de génération en génération

[1] L'adjectif *populaire* étant polysémique (à succès ou relatif au peuple), la chanson populaire est parfois aussi vue comme « celle qui appartient à tout le monde et à personne » (Duneton, 1998 : 241).

aussi bien oralement que par écrit, aussi bien populaire que littéraire » (Laforte, 1976 : 6) avaient pour leur part des acceptions trop larges. Une chanson de tradition orale ou chanson folklorique s'oppose selon Laforte (1976 : 7) à une chanson écrite selon les critères suivants : auteur anonyme/auteur connu, dimensions spatio-temporelles indéterminées/dimensions spatio-temporelles déterminées, composition orale/composition écrite, poétique inconnue/poétique connue et texte multiple/texte unique.

J'ai commencé à travailler pour mon mémoire de master 1 (Coutanson, 2016) sur des paroles de chansons écrites (issues de recueils ou de versions en ligne). Comme nous l'avons vu en 2.1.1, je travaillais à partir de collections d'exemples, en amassant des occurrences de pataquès. Si ceux-ci étaient systématiquement glanés dans les recueils, ils ne l'étaient pas sur les sites de chansons de tradition orale en raison du volume de titres à disposition. Doutant de la véracité des pataquès présentés, j'ai voulu travailler à partir de documents sonores. Comme je l'avais montré dans mon mémoire de master 2 (Coutanson, 2017 : 112), on peut noter de la variation dans la réalisation de pataquès pour un même contexte répété au fil d'une chanson, variation qu'un recueil ne présentera pas. On peut aussi douter qu'un pataquès soit un élément stable, systématiquement produit à chaque nouvelle **performance**[2]. Travailler sur des paroles écrites pourrait à la fois amener du bruit dans le corpus (pataquès relevés mais qui ne seraient pas systématiques, pataquès non produits par le chanteur mais ajoutés lors de la retranscription pour donner un caractère « populaire » à la chanson) ou du silence (pataquès produits mais non relevés par le transcripteur). Dans les recueils, les chansons sont également retravaillées : recoupement de diverses versions, correction des problèmes de métrique, raccourcis de notation (par exemple, non-recopie des vers bissés ou copie d'une seule version du refrain en ce qui concerne les paroles et choix d'une seule notation musicale pour les refrains et les couplets, occultant toute variation)… Il semblait donc essentiel d'avoir accès à de l'oral, mais comme nous allons le voir, pas à n'importe quel type d'oral.

J'ai alors commencé à faire des recherches, toujours en quête de productions authentiques, captées dans des situations d'enregistrement les plus écologiques possibles. Par exemple, sont accessibles un bon nombre de reprises par des chanteurs ou groupes traditionnels contemporains. Il est possible de bien distinguer ces deux types de productions comme le note Laforte (1976 : 10) chez certains jeunes collecteurs français de son

[2]La notion de *performance* a été « récupérée de la dramaturgie par l'ethnologie anglo-saxonne d'après-guerre » (Surmont, 2010 : 26). Considérer une chanson dans sa performance revient à prendre en compte « non seulement la vocalisation et la gestualité […] mais aussi l'interaction entre l'interprète, la scène et le public, l'éclairage […] », la tenue vestimentaire des chanteurs, etc… (Surmont, 2010 : 25-26, 30).

époque, qui utilisaient le terme ***chansons de collectage*** pour distinguer les chansons recueillies auprès de chanteurs traditionnels de ces chansons reprises. Ces dernières ne me semblaient pas adaptées, car l'écrit a pu biaiser au passage ces chansons, à l'encontre de la transmission orale qui avait cours jusqu'alors. Autrefois, une plus grande place était accordée à la variation dans ces formes de littérature orale (Duneton, 1998 : 241), tandis qu'aujourd'hui la plupart des chanteurs essayent d'être les plus fidèles possible à une version de référence. J'avais également travaillé sur des reprises de chanteurs de l'industrie du disque dans mon mémoire de master 2 (Coutanson, 2017 : 116-117), comme présenté en 1.3.6.1. Il se peut que lors de reprises de chants de tradition orale, une attention toute particulière soit accordée au maintien des pataquès, considérés comme caractéristiques de ce répertoire. Il n'en reste pas moins que le choix de travailler directement sur des chansons de tradition orale semblait plus aller de soi.

J'ai alors découvert tout d'abord le Portail du Patrimoine Oral[3], puis celui de la Base inter-régionale du patrimoine oral (BIPO)[4], qui est une « banque de données [qui] donne accès aux collections d'archives sonores du patrimoine oral de six régions : Normandie, Bourgogne, Limousin, Auvergne, Rhône-Alpes, Languedoc-Roussillon »[5], notamment à plus de 400h d'enregistrements inédits (BIPO, 2022a). La BIPO est alimentée par des structures associatives et territoriales (BIPO, 2022a) qui ont édité un guide des bonnes pratiques en 2014 afin de faire converger les pratiques en matière de traitement d'archives sonores inédites (Marcadé et al., 2014). Outre les chansons, la BIPO présente également des enregistrements de musiques traditionnelles, des contes, des récits de vie ou encore des paysages sonores ainsi que des productions de chanteurs professionnels, enregistrés à des époques différentes. Les données de la BIPO sont aussi présentes sur le Portail du Patrimoine Oral qui dépend de la Fédération des associations de musiques et danses traditionnelles (FAMDT), un catalogue réalisé avec le soutien de la BnF. C'est dans la BIPO que j'ai fait la rencontre du fonds Jean Dumas (dorénavant FJD). Quelques chansons utilisées dans les collections d'exemple de mes mémoires provenaient d'autres fonds, mais ce fonds-ci a été tout particulièrement attirant pour moi.

J'aurais encore pu décider de constituer par moi-même un nouveau corpus de chants de tradition orale. En master, j'ai même fait un stage à ce sujet au Centre des musiques traditionnelles en Rhône-Alpes (CMTRA). Outre le fait que je n'avais pas reçu de forma-

[3]Le Portail du Patrimoine oral a été mis en ligne en 2011 (Marcadé et al., 2014). Il est accessible à l'adresse suivante : http ://stq4s52k.es-02.live-paas.net/

[4]La BIPO est accessible à l'adresse suivante :
http ://patrimoine-oral.org/dyn/portal/index.seam ?page = home

[5]En 2015, une réforme territoriale a redéfini les régions françaises : il s'agit de la loi Nouvelle organisation territoriale de la République (NOTRe), promulguée le 7 août 2015 (Gouvernement, 2021).

tion en ethnomusicologie par ailleurs, cette entreprise ne m'a pas semblée nécessaire et pertinente étant donné la masse de données disponibles à l'heure actuelle dans les archives sonores qui ont déjà pu être numérisées et qui sont de fait dorénavant exploitables par les chercheurs. Enfin, collecter aujourd'hui un corpus de chansons de tradition orale me paraissait risqué pour deux raisons : tout d'abord car il existe une frontière poreuse entre le registre traditionnel et les chansons issues de l'industrie du disque, parfois intégrées au répertoire sans pouvoir aisément distinguer ce qui est là depuis plusieurs générations de ce qui a été récemment intégré, ensuite, car certaines chansons de tradition orale sont arrivées dans le répertoire via des recueils de chansons écrits, et donc sans transmission orale directe comme vu précédemment, ou via des chanteurs de l'industrie du disque (ce qu'on pourrait appeler la « génération Henri Dès ») avec à nouveau une transmission orale différente des canaux traditionnels. Je ne nie pas que cette façon de délimiter quels chants sont à considérer comme relevant de la tradition orale est problématique, mais je laisse cette prise de risque aux collecteurs, qui partagent ces réserves et qu'ils ont pu, par leur formation (études en ethnomusicologie) ou leur expérience de terrain, surpasser.

Je clôturerai cette partie avec quelques questions éthiques que posent l'existence de bases de données ethnomusicologiques libres d'accès, notamment en ce qui concerne la réutilisation commerciale de certains enregistrements issus de ce champ disciplinaire (Zemp, 1996). Si l'article de Zemp (1996) concernait la réutilisation de chants et musiques traditionnels provenant d'enquêtes de terrain et diffusés sur des cassettes et CDs, cette question est renouvellée par le dépôt grandissant de ces enregistrements sur des bases de données libres d'accès. La réinterprétation de ces œuvres musicales n'est pas problématique, mais la réutilisation des œuvres originales l'est, par exemple sous forme de *samples*, bien qu'elles soient souvent considérées comme relevant du domaine public. Zemp (1996) mentionne notamment l'exemple bien connu d'un chant de tradition orale mélanésien (interprété par Afunakwa, du peuple baegu de Malaita, dans les Îles Salomon), réutilisé dans l'album *Deep Forest* du groupe éponyme, composé des musiciens français Michel Sanchez et Eric Mouquet, sous le nom de "Sweet Lullaby", titre qui s'est hissé au top 50 français et s'est maintenu de nombreuses semaines en tête des charts américains, tout en servant de musique à une publicité de la marque de produits d'hygiène Ushuaïa. Les musiciens n'ont jamais contractualisé la réutilisation de l'œuvre et n'ont reversé de dividendes ni à la chanteuse, ni à une association ou un musée local, ni à l'ethnomusicologue ayant collecté le chant traditionnel, ni à la maison d'édition ayant publié la version originale, ce qui a causé un scandale vu le succès du titre et les profits potentiellement engendrés. À l'heure de la science ouverte, nous devons être

conscients de ces réutilisations problématiques des œuvres que nous mettons à disposition du public.

En résumé, après avoir opté pour l'appellation *chansons de tradition orale* pour se référer à mon objet-chanson, j'ai expliqué comment, pour mes mémoires de master, je suis passée d'une collection d'exemples portant sur des paroles de chansons écrites à une collection d'exemples oraux afin d'éviter le bruit et le silence générés par les corpus écrits. J'ai aussi privilégié des données ethnomusicologiques pour être au plus près de chaque performance des chanteurs. J'ai enfin entrepris d'exploiter la richesse de la BIPO en travaillant sur le FJD mais écarté la possibilité d'enregistrer moi-même un nouveau corpus de chants de tradition orale en raison de l'abondance de tels corpus numérisés et de l'interférence intensifiée de recueils écrits ou de chansons issues de l'industrie du disque sur le registre traditionnel des chanteurs contemporains. Pour finir, j'ai présenté une question éthique soulevée par la mise à disposition de chants traditionnels dans des bases de données libres d'accès. Poursuivons. Qui dit FJD dit Jean Dumas : qui était-il ? Quelles étaient ses motivations ? De quoi est composé le FJD ?

3.2.2 Jean Dumas, folkloriste à l'origine du fonds Jean Dumas

FIGURE 3.1 : Virginie Granouillet et Jean Dumas en 1960

Les grands-parents de Jean Dumas (1924-1979) ont vécu à Vorey-sur-Arzon, dans le département de la Haute-Loire, où son grand-père tenait un bureau de tabac (Desgrugil-

lers, 2014 : 7,35). Lui-même est né à Clermont-Ferrand, fils unique d'un père cheminot et d'une mère buraliste (Rocher, 1994 : 35). Agrégé d'italien, il enseigne à Moulins puis à la faculté de lettres de Clermont-Ferrand (Desgrugillers, 2014 : 7).

Enfant, Jean a été Éclaireur, puis a été très impliqué dans le mouvement ajiste (auberges de jeunesse ; Rocher, 1994 : 36). Il aime la musique : il pratique lui-même la guitare, la flûte à bec ainsi que le chant et enseigne la musique en plus de l'italien (Rocher, 1994 : 36). Il a également été chef de chœur et publie des recueils de chant (Rocher, 1994 : 36, 40). Déçu par les ouvrages de l'époque, il cherche très tôt à collecter lui-même des chansons afin de renouveler le répertoire existant (Rocher, 1994 : 37, 39). En fin de vie, il préparait aussi un projet de thèse portant sur le livret d'opéra (Rocher, 1994 : Annexe 2, 163).

Les chansons du FJD dont il est question dans cette thèse ont été enregistrées de 1958 à 1962 (Tableau 3.1), peu après l'acquisition en 1956 d'un premier magnétophone par notre folkloriste (Rocher, 1994 : 46). En 1959, il se marie avec Thelma, assistante d'anglais à Moulins d'origine galloise, qui l'assistera dans sa collecte (Rocher, 1994 : 63-64 ; Jaulhac, 2013). Notons que Jean et Thelma ont eu un premier enfant en 1961 (Rocher, 1994 : 69). En tout, Jean aura effectué des collectes de 1945 à 1972 (Rocher, 1994 : Annexe 3, p.167). Sur la photographie 3.1, on retrouve Jean Dumas aux côtés de Virginie Granouillet, aussi appelée La Baracande[6], la chanteuse au répertoire le plus étendu du FJD (un des corpus du fonds lui est consacré, voir tableau 3.1) qui impressionnait beaucoup notre collecteur, avec ses 178 chansons enregistrées en onze séances (Rocher, 1994 : 52, 54).

Le FJD, accessible sur le site de la BIPO, est en phase d'être subdivisé en cinq sous-fonds (BIPO, 2022f) :

- JD_a, ou collectes sonores de Jean Dumas
- JD_b, ou fiches des chansons rédigées par Jean Dumas
- JD_c, ou articles - écrits de Jean Dumas
- JD_d, ou photos prises en collecte
- JD_e, ou correspondance de Jean Dumas.

C'est ce qui est annoncé sur le site de la BIPO, bien que pour le moment seules des archives sonores et des fiches accompagnant les chansons aient été mises en ligne. En ce qui concerne ces éléments, le FJD présente actuellement cinq corpus (Tableau 3.1 ; BIPO, 2022b,c,d,g,h). Le cinquième corpus, *Chansons et danses de la Montagne Bourbonnaise : les enquêtes de Jean Dumas dans l'Allier (03)*, mis en ligne en 2021, n'a pas été

[6]Desgrugillers (2014 : 7) nous apprend qu'il s'agit d'un *nom d'escaine*.

pris en compte dans le cadre de ce travail de recherche.

Nom du corpus	**Date de mise en ligne**	**Lieu d'enregistrement**	**Date d'enregistrement**
Les chansons de la Baracande : les enquêtes de Jean Dumas auprès de Virginie Granouillet	2010 et 2011	Mans (Roche-en-Régnier, Haute-Loire)	1958-1962
Voix de l'Emblavez et du Velay : les enquêtes de Jean Dumas dans la Haute-Loire	à partir de 2011	Emblavez (Haute-Loire)	1958-1962
Chansons de couteliers : les enquêtes de Jean Dumas dans les Bois Noirs et la vallée de la Durolle	à partir de 2014	Montagne thiernoise, vallée de la Durolle (Puy-de-Dôme)	1958-1961
Chansons de dentellières : les enquêtes de Jean Dumas auprès des dentellières de la vallée de la Loire	2014	vallée de la Loire (Haute-Loire)	1958-1962
Chansons et danses de la Montagne Bourbonnaise : les enquêtes de Jean Dumas dans l'Allier (03)	2021	Doyet, montagne bourbonnaise (Allier)	1960-1961

TABLEAU 3.1 : Corpus compris dans le fonds Jean Dumas

Dans le cadre de cette thèse, j'étudie un échantillon de 498 chansons issues des quatre premiers corpus du FJD susmentionnés, d'une durée de 18h 23min 30s, comprenant 89 853 mots[7] pour 5 787 contextes de liaison possibles.

Pour conclure brièvement sur le personnage de Jean Dumas, celui-ci a collecté ce qu'on nomme aujourd'hui le *fonds Jean Dumas* (FJD) entre 1958 et 1962 dans le centre de la France, où une partie de sa famille avait habité. La musique était un élément important dans sa vie, que ce soit sur le plan professionnel ou dans son temps libre. Le FJD, dont la mise en ligne n'est pas terminée, devrait comporter cinq sous-fonds. Le sous-fonds sonore est présentement constitué de cinq corpus, dont un de couteliers, un de dentellières et un centré sur la Baracande, une chanteuse traditionnelle prolifique. Maintenant que nous avons fait plus ample connaissance avec Jean Dumas et le FJD, nous allons nous intéresser à la démarche de ce collecteur lors de son recueil, et c'est par les forces de cette démarche que nous débuterons notre observation.

[7] Comme nous l'avions indiqué dans un article (Coutanson & Badin, 2021 : 136), « [l]e nombre de mots dans le corpus a été calculé avec la commande FREQ du logiciel CLAN qui permet d'éviter de prendre en compte la ponctuation et surtout les codes propres au format CHAT ». Ici, j'ai additionné le nombre de tokens pour SGR (89 559), SGR1 (151) et SGR2 (143).

3.2.3 Démarche de Jean Dumas

3.2.3.1 Forces de la démarche de Jean Dumas

Jean Dumas remet en cause la démarche des spécialistes de son époque, et notamment de Patrice Coirault. Marlène Belly, ethnomusicologue du domaine francophone, nous rappelle que ce dernier est « l'auteur d'une œuvre magistrale pour l'ethnomusicologie francophone » (Belly, 2015 : 566). Coirault travaillait à partir d'une immense collection de documents écrits, répartis en différentes bases de données, aujourd'hui léguée à la Bibliothèque nationale de France et qu'il avait lui-même constituée. L'ensemble n'était pas informatisé mais était relié à un fichier central comprenant d'innombrables fiches interconnectées les unes aux autres (Belly, 2015 : 567). Marlène Belly nous apprend que les travaux de Coirault « se sont [...] orientés vers la recherche des lignages et centrés autour du concept de chanson-type » (Belly, 2015 : 568). Luc Charles-Dominique, lui aussi ethnomusicologue, précise également que Coirault « s'intéresse avant tout aux processus de réélaboration permanente liés à la transmission orale » (Charles-Dominique, 2007). S'il accorde une attention particulière à la variation, c'est parce qu'il y cherche une certaine systématicité car il a entrepris de « réaliser une classification générale des chansons populaires » (Charles-Dominique, 2007). Bien que Coirault ait lui-même fait du terrain une cinquantaine d'années auparavant (Charles-Dominique, 2007), l'éminent chercheur assure qu'il est trop tard pour collecter des chants de tradition orale :

> [l]'époque opportune d'une collecte profitable [...] est passée. Après s'être tant bien que mal prolongée jusqu'aux approches de la première guerre mondiale, elle s'est alors définitivement close. [...] Vainement des discophiles en tournée s'inquiètent [des vieilles chansons] [...], afin de leur courir sus, armés de leurs phonos. [...] Cependant, les documents transcrits ou publiés aux époques favorables les attendent dans les bibliothèques. [...] Quels qu'ils soient, ils détiennent la connaissance de notre ancienne chanson folklorique ; ils la détiennent seuls ; on ne l'acquiert plus que par leur intermédiaire (cité par Etay dans Rocher, 1994 : Annexe 1, p.158-159).

De cette façon, ayant pourtant été lui-même en rupture avec ses prédécesseurs et certains de ses contemporains dont la quête de l'origine des chansons populaires lui semblait vaine (Charles-Dominique, 2007), Coirault crée à son tour une rupture avec certains de ses successeurs en désapprouvant la démarche de collecte tardive des folkloristes qui l'ont suivi et leur appui sur l'enregistrement sonore.

Jean Dumas, nonobstant les recommandations de Patrice Coirault, comprend, et

c'est là une deuxième force, qu'il y a urgence à poursuivre les collectes, d'autant que les témoins qu'il enregistre sont déjà assez âgés, car elles pourraient se révéler encore fructueuses. Notre collecteur est un passionné, Claude Rocher parle de « réelle boulimie » (Rocher, 1994 : 54), ce qui lui permet de se lancer ardemment dans ce recueil. Nous avions vu plus haut (3.2.2) que les chansons du FJD avaient été enregistrées de 1958 à 1962. Pour Virginie Granouillet, morte en 1962 (Rocher, 1994 : 69), la collecte a été faite à point nommé, et j'avais déjà mentionné à quel point celle-ci avait été prolifique. Il y a ici un parallèle saisissant à faire avec la linguistique de terrain et le mythe des « derniers locuteurs » (Bert & Grinevald, 2010 : 129). Ici, les folkloristes se dépêchent d'enregistrer les derniers chanteurs traditionnels. Cette dimension de conservation de la parole au-delà de la mort existe depuis les débuts de l'enregistrement sonore (Cordereix, 2014 : 10). Jean Dumas se tourne préférablement vers des chanteurs âgés, car c'est auprès d'eux qu'il pense trouver de l'authentique. Cette approche est semblable à celle des linguistes de terrain qui pensent trouver de l'authenticité auprès des locuteurs traditionnels âgés ou d'anciens locuteurs (Bert & Grinevald, 2010 : 125,127).

Pour continuer, c'est une période de grands changements dans les lieux d'enregistrement de notre collecteur, période dans laquelle s'inscrit cette pratique des chants de tradition orale qui est amenée à être redéfinie : la population est en constante diminution depuis la fin du XIX^{ème} siècle suite à des vagues d'émigration ayant fourni de la main d'œuvre au secteur minier, ayant participé à la construction des ouvrages d'art pour le chemin de fer ou encore étant partie renforcer les rangs des travailleurs pour les gros travaux agricoles des régions voisines (Rocher, 1994 : 16, 18). Le secteur dentellier est pour sa part en perte de vitesse suite à la mécanisation (Rocher, 1994 : 27). Cela fait une dizaine d'années que les automobiles ont fait leur apparition dans la région après le chemin de fer à la fin du XIX^{ème} siècle, l'eau courante et l'électricité se font plus répandues bien que toutes les maisons ne soient pas encore raccordées, certaines d'entre elles deviennent des résidences secondaires, les premiers vacanciers arrivent, le nouveau franc est introduit en 1960 ainsi que les premiers transistors (Rocher, 1994 : 18, 20, 22, 60)... On sait également que certains chanteurs ne savaient ni lire ni écrire (comme Virginie Granouillet, cf JD_a062_13, 95), quand d'autres ont été alphabétisés (comme Henriette Usson, cf JD_a330_01, 95). On pourrait penser que depuis les Lois Ferry, notamment celle de 1882 rendant l'école obligatoire, tous les enfants français sont scolarisés. Virginie Granouillet, née en 1878, nous démontre le contraire : cela fait que dans les années 1960, il est encore possible d'enregistrer des locuteurs/chanteurs n'ayant pas reçu d'instruction.

Durant cette période de grands changements, certains éléments facilitent l'entre-

prise de Jean Dumas. Pour mener à bien sa collecte, il dispose notamment d'une automobile (Photographie 3.2) qui lui permet d'accéder à ses terrains, qui sont assez reculés, et bénéficie de routes alors praticables, ce qui constitue une véritable révolution dans le monde des collecteurs de chants de tradition orale (Rocher, 1994 : 20, 43). En outre, les technologies d'enregistrement sonore deviennent plus accessibles et mobiles. La captation du FJD est rendue possible par l'acquisition d'un premier magnétophone par notre collecteur en 1956, un Lugavox, puis d'un Stuzzi en 1959 par Thelma, pour lesquels les époux se fournissent en bandes magnétiques (Rocher, 1994 : 46, 51). Par ailleurs, ce professeur de lycée dispose de temps libre (vacances scolaires, jours fériés) pour effectuer ses collectes (Rocher, 1994 : 54). Enfin, Jean peut compter sur la collaboration de son épouse Thelma dès leur mariage en 1959 (Rocher, 1994 : 63-64).

Contrairement à ses prédécesseurs, Jean Dumas cherche de l'authentique et ne souhaite pas d'intermédiaires, ce que lui permet l'accès à son terrain par ses propres moyens, comme nous venons de le voir (au contraire le collecteur Victor Smith, qui a effectué des collectes dans la même région à une époque où elle était encore plus difficile d'accès, pouvait recevoir des paroles de chants de collaborateurs ; Rocher, 1994 : 110-112). En outre, le FJD a cela d'authentique que les chansons ne sont généralement pas préparées avant la performance, que les chanteurs ont été présentés à notre collecteur par des connaissances (Jean Dumas utilise la **méthode boule de neige** (Rocher, 1994 : 46) qui consiste à partir de son réseau personnel et à rencontrer de nouveaux locuteurs en demandant à ses connaissances de présenter à leur tour de nouveaux locuteurs qui seraient adaptés pour la tâche linguistique à effectuer ; dans notre cas, il s'agirait de pouvoir interpréter des chants de tradition orale), qu'ils sont dans un lieu qui leur est familier, à leur domicile ou chez une connaissance, lieu qui peut être leur lieu habituel pour cette activité de chant, par exemple pour les dentellières. Pour ces dernières, si le lieu et le public sont habituels, les conditions de chant peuvent différer (une seule chanteuse par exemple), ainsi que l'activité sociale dans laquelle ces chants s'inscrivent. Si la plupart des chansons du sous-corpus de dentellières semblent enregistrées hors contexte de couvige, Jean Dumas a tout de même essayé de faire quelques prises au son des fuseaux (cf photographie 3.3). Pour d'autres témoins, cela ne pouvait pas être mis en place, notamment pour les scieurs de long (trop bruyant, nécessité de trouver des scieurs de long encore en activité) ou pour les chansons de quête, comme pour les **chants de mai** (mauvaise période de l'année, nécessité de trouver des groupes de jeunes gens faisant perdurer cette tradition). Le blog du CMTRA nous renseigne sur cette pratique :

> Dans la nuit [du] 30 avril au 1er mai, des groupes de jeunes gens parcouraient les chemins, s'arrêtant à toutes les maisons pour y chanter le mai. Les chanteurs ne s'arrêtaient de chanter que lorsque les habitants descendaient pour remplir le panier d'œufs et de lards, qui servaient le dimanche suivant à faire une grande omelette pour les jeunes du village. [...] D'autres témoignages laissent cependant penser que, au delà de la célébration du printemps (qui s'inscrit dans une tradition païenne très ancienne et répandue en Europe), les quêtes de mai servaient également à faire se rencontrer les jeunes des hameaux voisins, et jouait vraisemblablement un rôle dans le marché matrimonial local (CMTRA, 2017).

Enfin, Jean s'intéresse au mode de transmission des chansons qu'il recueille, afin de garantir à nouveau leur authenticité. Il demande parfois aux individus qu'il enregistre d'où proviennent leurs chansons. Sont cités les parents (par exemple, Marie Rigodias indique en JD_a168_04, 18 tenir la chanson de sa mère), d'autres villageois ou encore des bourgeois (comme Henriette Usson, cf JD_a330_01, 101-104) ou des fermiers chez qui les enfants pouvaient être placés. Jean-Pierre Arnaud a aussi indiqué avoir chanté en labourant (JD_a163_14, 49-50) et Marie Fontenille (qui exerçait le métier de coutelière) à l'usine (JD_a168_14, 120). Enfin, un chanteur a mentionné que les chansons étaient occasionnellement présentées dans des foires.

Dans cette quête d'authenticité, c'est surtout sa décision d'enregistrer les chanteurs, favorisée comme nous l'avons vu par l'acquisition de magnétophones, qui constituera le plus bel atout de ce corpus, comme le note Françoise Etay, responsable du Département de Musique Traditionnelle du Conservatoire de Limoges : « Il est donc probable que ce sont ses enregistrements, plus que ses écrits, qui resteront le plus beau cadeau qu'ait laissé Jean Dumas aux générations futures » (Etay dans Rocher, 1994 : Annexe 1, p.161). S'ensuit une possible réécoute par notre collecteur lors de la rédaction des fiches accompagnant les audios, documents sonores qu'il pouvait également faire écouter à d'autres, et qui sont nécessaires pour étudier ces deux phénomènes oraux que sont le pataquès et la liaison. Un accès aux données primaires a constitué une condition *sine qua non* lors de ma sélection d'un corpus. Mais quelle que soit notre discipline, leur existence permet de vérifier les données secondaires rédigées par Jean Dumas, qui ne peuvent être tout à fait fidèles aux productions linguistiques et musicales des chanteurs. Par exemple, au niveau linguistique, Jean Dumas ne retranscrit pas les disfluences, il opte parfois pour un des vers lorsqu'un passage d'une chanson est censé être répété mais qu'il y a de la variation ou encore malgré l'utilisation d'une orthographe adaptée,

FIGURE 3.2 : Jean Dumas et sa 2 CV

FIGURE 3.3 : Séance d'enregistrement des dentellières par Thelma

comme nous le verrons par la suite, il n'indique pas toutes les productions phonétiques non standard. Sur le plan musicologique, les audios se révèlent aujourd'hui essentiels car d'après Françoise Etay, les notations de Jean Dumas laisseraient à désirer : il faisait fi des variations musicales ténues, se serait peu relu et n'utilisait que des mesures très simples, quelquefois inadaptées (Etay dans Rocher, 1994 : Annexe 1, p.160). Il faut savoir que l'utilisation de la notation musicale occidentale a été décriée en ethnomusicologie, notamment les concepts de *note*, de *hauteur*, de *mesure*, car elle ne permet pas de rendre compte de toutes les productions musicales humaines, notamment pour la musique vocale de tradition orale et les musiques non occidentales (cf par exemple Will, 1999). Comme le résume Will (1999), « c'est probablement l'enregistrement qui nous permet de nous rapprocher le plus de la musique si nous devons choisir un moyen pour la conserver ». Par exemple, le FJD comporterait des exemples d'**aksaks**[8], ce qui semble n'avoir jusqu'alors jamais été observé dans la musique de tradition orale française et il aurait été impossible d'en relever d'après les transcriptions de Jean Dumas

[8]Il s'agit d'un « système rythmique au sein duquel des pièces ou des séquences se déroulant généralement dans un tempo vif, reposent sur la réitération ininterrompue d'un module résultant de la juxtaposition de groupements fondés sur des quantités binaires et ternaires (telles que 2+3, 2+2+3, etc.) et dont le nombre global est le plus souvent impair » (Arom, 2004).

seules (Etay dans Rocher, 1994 : Annexe 1, p.160). On voit tout l'intérêt de conserver des données primaires, pour révéler *a posteriori* des éléments absents des données secondaires produites par Jean Dumas. Pourtant, Will (1999) nous rappelle que :

> [l]es enregistrements partagent avec les notations un point commun : comme tous les systèmes d'écriture et de notation musicales, l'enregistrement n'est pas à même de fixer et de restituer tous les aspects de ce qui a été dit, chanté ou joué. Il ne parvient à rendre ni le contexte dans lequel un événement musical a eu lieu, ni sa signification, ni les indications sur l'attitude du musicien par rapport à ce qui a été enregistré.

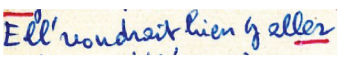
Ce commentaire est tout autant pertinent pour un linguiste, car le contexte de production des chansons peut influencer la réalisation ou non de certains éléments linguistiques, et ce contexte ne nous est pas entièrement connu à l'écoute des documents sonores. Par ailleurs, comme l'indiquent Baude et al. (2006 : 46),

> [...] tout enregistrement est le fruit de décisions à la fois techniques et théoriques – concernant par ex. le choix du moment à enregistrer et la délimitation du segment enregistré, le choix [...] du positionnement et de l'orientation du micro pour l'audio – qui reposent sur une connaissance préalable de l'activité enregistrée.

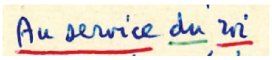
En ce qui concerne Jean Dumas, celui-ci ne positionne pas forcément le micro de façon à rendre au mieux la performance des chanteurs, ceux-ci chantant parfois très faiblement. En outre, l'utilisation d'un micro unique, qui plus est sans enregistrement vidéo, ne permet pas de comprendre toutes les interventions du public et de savoir combien d'auditeurs interviennent. Pour finir, notre collecteur n'actionne pas toujours l'enregistrement sonore au bon moment, soit parce qu'il laisse un premier temps aux chanteurs pour se remémorer la chanson, soit parce qu'il évalue si la chanson proposée a déjà été enregistrée. Cela éloigne les chanteurs de performances réelles, car il leur demande souvent de recommencer au début, les fait reprendre s'ils sautent une strophe ou les pousse à se remémorer la fin de chansons.

Les fiches rédigées par Jean Dumas accompagnant la majeure partie des enregistrements se sont également révélées très utiles afin de vérifier mes propres transcriptions, leur précision phonétique et les informations sociologiques sur les chanteurs interrogés associées auront elles aussi tenu une part non négligeable dans mon choix de ce corpus. Claude Rocher note la « rigueur extrême de Jean qui le dissuade de jamais corriger le texte [...] » (Rocher, 1994 : 81), rigueur qui le poussera jusqu'à utiliser parfois

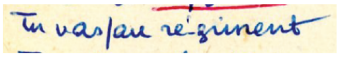
une orthographe adaptée. S'il n'est pas linguiste, il présente toutefois une conscience linguistique assez développée, notamment au niveau phonétique où il annote, le plus souvent avec justesse, des (non)-réalisations de schwas (voir illustrations 3.4a et 3.4b) ou de liaisons (voir illustrations 3.4c et 3.4d), des pataquès (voir illustration 3.4e), des changements (voir illustrations 3.4f et 3.4g) ou harmonies vocaliques (voir illustration 3.4h), des non-réalisations de la consonne finale du pronom sujet *il* devant initiale consonantique (voir illustration 3.4i), des simplifications de groupes consonantiques (voir illustration 3.4j[9]) ou encore des réductions (voir illustration 3.4k). Ses notations



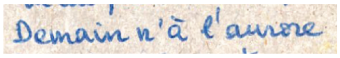
(a) Non-réalisation d'un schwa : *Ell' voudrait bien y aller* [ɛlvudrɛbjɛ̃jale], extraite de la fiche associée à JD_a061_14



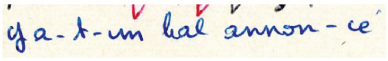
(b) Réalisation d'un schwa : *Au service du roi* [oseʀvisə-dyrwa], extraite de la fiche associée à JD_a061_13



(c) Non-réalisation d'une liaison : *Tu vas|au ré-giment* [tyvaoreʒimɑ̃], extraite de la fiche associée à JD_a061_13

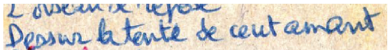


(d) Réalisation d'une liaison : *Demain n'à l'aurore* [dəmɛ̃nalororə], extraite de la fiche associée à JD_a061_04



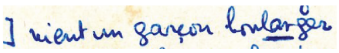
(e) Réalisation d'un pataquès : *Y a-t-un bal annon-cé* [jatœbalanɔ̃se], extraite de la fiche associée à JD_a061_14

(f) Changement vocalique : *Au baur de Nantes* [obordənɑ̃tə] avec [bor] et non [bɔr] pour *bord*, extrait de la fiche associée à JD_a061_14



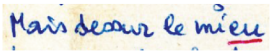
(g) Changement vocalique : *Dessur la tente de ceut amant* [dəsyʀlatɑ̃tədəsətamɑ̃] avec [səʔ] et non [sɛt] pour *cet*, extrait de la fiche associée à JD_a061_04

(h) Harmonie vocalique : *Le minuisier habile* [ləminɥiz-jeabilə] au lieu de *menuisier* [mənɥizje], extraite de la fiche associée à JD_a060_24

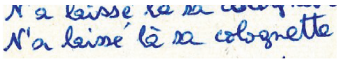


(i) Non-réalisation du [l] de *il* devant initiale consonantique : *I vient un garçon boulanger* [ivjɛ̃tœ̃gaʀsɔ̃bulɑ̃ʒe], extraite de la fiche associée à JD_a061_01

(j) Simplification de groupe consonantique : *A not' supérieuse.* [anotsyperijøzə] au lieu de *notre* [notʀ], extraite de la fiche associée à JD_a060_24



(k) Archaïsme et réduction de [lj] à [j] : *Mais dessur le mieu* [medəsyʀləmijø] au lieu de *sur* et *milieu* [miljø], extraits de la fiche associée à JD_a061_03



(l) Terme rare : *N'a laissé là sa colognette*, extrait de la fiche associée à JD_a061_08

FIGURE 3.4 : Exemples de graphies non standard utilisées par Jean Dumas

[9]Notons au passage le féminin en *supérieuse* pour *supérieure* sur le modèle de *chanteur/chanteuse* pour rimer avec le vers précédent en *religieuse*.

des paroles se sont par ailleurs révélées très utiles lorsque l'audio laissait à désirer ou encore en cas d'utilisation d'un terme rare (voir *colognette*[10] dans l'illustration 3.4l) ou d'archaïsmes (voir *dessur* dans l'illustration 3.4k) qui ne m'étaient pas connus. Qui plus est, Jean Dumas indique pour chaque chanteur ses nom et prénom, son âge, son lieu de résidence précis (village, hameau voire lieu-dit), son métier, la date de la collecte et quelquefois d'autres informations comme son surnom, sa date de naissance, le nom de naissance des femmes mariées... Toutes ces informations forment les métadonnées associées à chaque document sonore dans mon corpus, constitué des données primaires du FJD pour lesquelles j'ai créé des données secondaires (transcriptions et annotations pour la liaison, le pataquès et le schwa).

Jean Dumas, à contre-courant, ne cherche ainsi pas à établir une chanson de référence mais vise les performances uniques et authentiques de chanteurs traditionnels en faisant du terrain dans un environnement qui leur est familier. Il collecte avec empressement ces chants de tradition orale car les chanteurs, qui ne sont pas tous alphabétisés, sont âgés et que le monde qui les entoure évolue rapidement. Jean Dumas bénéficie d'avancées technologiques, de temps libre en dehors de son travail et de la collaboration de son épouse pour mener à bien son entreprise. La richesse du FJD vient avant tout des enregistrements faits par notre collecteur et conservés jusqu'à nos jours. Cet accès aux données primaires permet d'approcher au plus près les performances des chanteurs et de revoir les notations linguistiques et musicales de Jean Dumas (en dévoilant au passage d'improbables aksaks). Ses fiches, où il adopte une orthographe adaptée pour rendre compte de traits phonétiques non standard, m'ont permis un accès aux métadonnées des chansons et de vérifier mes propres transcriptions/annotations. Le FJD, aussi impressionnant soit-il, présente tout de même certaines faiblesses. J'en ai déjà brièvement mentionné dans cette partie, mais je vais à présent émettre les critiques majeures qui peuvent être faites à l'égard de la méthodologie de Jean Dumas.

3.2.3.2 Faiblesses de la démarche de Jean Dumas

En premier lieu, Jean Dumas accorde peu d'importance aux chansons en langues régionales. Dans notre cas, il semble le plus souvent qu'il ait été question d'occitan, langue qui est pour lui une langue d'héritage mais qu'il maîtrise peu (Rocher, 1994 : 44). Leur faible nombre dans le FJD nous fait nous demander si notre collecteur en a peu conservé,

[10] Pour Desgrugillers (2014 : 140), *colognette* serait un occitanisme désignant la *quenouillette*, de l'occitan *colonheta*.

s'il a orienté au préalable ses demandes lors de ses collectes pour les éviter, ou si les chanteurs en avaient peu dans leur répertoire. Jean Dumas voyait d'un mauvais oeil les traductions vers le français, qui pour lui dénaturaient soit le sens, soit l'esthétique des œuvres (Rocher, 1994 : 78). Il estimait qu'« on ferait mieux de faire chanter [ces chansons] telles quelles [...]. Il y a des choses trop dures pour être chantées en dialecte par nous. Il faut donc les laisser aux gens du pays » (Rocher, 1994 : 78). Le critère de la langue comme critère de délimitation du corpus qu'il souhaitait recueillir pourrait tout à fait être recevable, si ce n'était à nouveau un objet linguistique que l'on éloigne car on le maîtrise peu. Il ne faut pas oublier que l'objectif de Jean Dumas en collectant était de renouveler le répertoire francophone existant à destination des chanteurs (Rocher, 1994 : 37). Ici, son intérêt principal n'était pas ethnomusicologique ou de préservation d'éléments patrimoniaux menacés, auquel cas il aurait eu tout intérêt à conserver ces chansons en langues régionales, mais il paraît difficile de nier le poids d'une certaine glottophobie dans la sélection des éléments considérés d'intérêt pour le grand public, comportement qui n'est pas spécifique à notre collecteur à cette époque, dans un pays où les langues régionales ont longtemps été minorées face à une langue dominante dont on a une vision prestigieuse mais aveuglante. Notons pour clore ce sujet que Jean Dumas ne disposait que d'outils (liste d'incipit) permettant d'enclencher des chansons en français (Rocher, 1994 : 79), ce qui pourrait les avoir favorisées.

En second lieu, Jean Dumas est très économe (Rocher, 1994 : 51). Il préserve les bobines, mais comme le note Alain Savouret, compositeur, dans sa préface de l'ouvrage *En écoutant la Baracande : avant le grand silence...* : « [à] la décharge de Jean Dumas, on sait que la bande magnétique n'est pas un support courant dans les années 50, coûteux donc ; d'un autre côté, il semble considérer l'enregistrement au service de l'écriture scrupuleuse de ses fiches quadrichromes » (Savouret dans Rocher, 1994 : Préface, p.10). Par ailleurs, ce sont les chansons qui intéressent Jean Dumas, il conserve ainsi peu de productions autres : « tout celà est le plus souvent "coupé", de façon impitoyable. A chaud tout d'abord, lors des séances d'enregistrement ; puis encore vraisemblablement au moment du "repiquage", c'est-à-dire lorsque Jean Dumas regroupe [les chansons] sur une même série de bobines » (Rocher, 1994 : 52). Pour les linguistes, cela aurait été pourtant crucialement pertinent de pouvoir davantage comparer les productions de mêmes locuteurs/chanteurs selon qu'il soit question des modalités parlée ou chantée. Il économise aussi la surface écrite, ce qui l'amène à utiliser des signes de reprise sur la partition pour éviter de répéter le refrain ou un vers qui serait redoublé lors de la transcription des paroles ou de la mélodie (Etay dans Rocher, 1994 : Annexe 1, p.160). Nous avions vu en 3.2.1 qu'il s'agissait d'une pratique courante de présentation des

chants de tradition orale dans les recueils.

Par ailleurs, Jean Dumas est persuadé du bien-fondé de son entreprise et peut se comporter quelquefois de manière inadaptée afin de parvenir à ses fins : il presse ses témoins pour que les chansons rentrent sur la fin de la bobine, leur demandant parfois de déclamer leurs chansons (quitte à ne pas les redoubler) au lieu de les chanter pour gagner du temps (exemple 125), ce qui n'est pas naturel pour eux ; il les pousse à chanter même lorsque l'envie n'y est pas ou quand ils n'ont pas le temps (exemple 126), à poursuivre car il a connaissance de couplets supplémentaires suite à ses enregistrements auprès d'autres enquêtés, et il lui est encore arrivé de faire croire aux chanteurs que le magnétophone était éteint alors que ce dernier enregistrait (exemple 127).

(125) *SGR : <la jeun(e) Hélèn(e)> [= ! sings].
*JD : non en en parlant en parlant maint(e)nant la jeun(e) Hélèn(e).
*SGR : <la jeun(e) Hélèn(e)> [= ! sings].
*JD : en parlant.
*SGR : en parl(ant) la jeun(e) Hélèn(e) ell(e) voudrait bien y aller.
(JD_a043_22, 43810, Marie Soulier)

(126) *SGR : oh j'ai pas le temps vous comprenez moi je suis pressée.
*JD : voil(e)s au vent.
*AUD : et oui mais qu'est-ce que vous voulez...
(JD_a359_02, 43800, Mme Rostin)

(127) *SGR : ç(a) ça marque là hein.
*SGR : xxx.
*JD : non non, non non.
*JD : non non.
(JD_a359_05, 43800, Mme Rostin)

Son éthique des relations interpersonnelles (Blanchet, 2000) laisse de la sorte parfois à désirer, privilégiant sa collecte, et cela renforce le **paradoxe de l'observateur**[11], au risque de déconnecter encore plus la performance en cours des performances habituelles, hors de toute observation par un individu extérieur à la communauté de pratiques. De même, Jean Dumas fait occasionnellement l'erreur d'utiliser des termes techniques que ne comprennent pas les chanteurs et qui pourraient les mettre en insécurité :

[11] Blanchet (2000 : 47) décrit le phénomène de la sorte : « la présence de l'enquêteur modifie les pratiques qu'il souhaite observer ».

(128) *JD : Alors c'est une complainte ça ?
*SGR : Ouh chais pas moi, j'le comprends pas.
(JD_a165_15, 43800, Germaine Varenne)

Enfin, les chanteurs n'ont pas forcément l'habitude d'être enregistrés et peuvent être perturbés par une technologie qui ne leur est pas familière :

(129) *SGR : Mais, i(l) faut mettre c(e) machin chez moi ?
*AUD : Non non, xxx chantez rien qu(e) !
'Non non, chantez seulement'
(JD_a354_02, 43800, Pauline Chazal)

Il faut noter en outre que Jean Dumas a peu connaissance d'entreprises similaires en son temps. Il connaît les travaux de Julien Tiersot, d'Arnold Van Gennep, de Patrice Coirault et correspond avec un collègue italien, Emilio Tron, qu'il finira par rencontrer (Rocher, 1994 : 38, 63). On sait à quel point l'accès à l'information est à l'époque différent, mais aussi perfectionnée que soit sa méthodologie voire en avance sur son temps, cette démarche solitaire est criticable. Jean Dumas a pourtant essayé de défendre son approche. Il a notamment écrit à William Lemit, Instructeur National de Chant au Ministère de la Jeunesse et des Sports et aux Centres d'entraînement aux méthodes d'éducation active (CEMÉA), mais a reçu une réponse pour le moins sèche en 1960 :

> Je ne crois nullement pour ma part à la valeur de procédés qui ont été proposés dans l'intention - parfaitement vaine à mon avis, et du reste sans intérêt - de rendre compte de toutes les imprécisions d'intonation et de rythme qui sont le fait courant des informateurs, surtout quand ceux-ci sont relativement âgés [...] A mon avis une notation honnête doit, à travers ces variations, s'efforcer de déceler les intentions du chanteur au-delà des faiblesses de sa réalisation (Etay dans Rocher, 1994 : Annexe 1, p.158),

avec cette fois-ci un parallèle saisissant avec le cadre génératiViste en linguistique et la compétence primant sur la performance. Françoise Etay note :

> On comprend mieux maintenant qu'à la fin des années cinquante, les collectes enregistrées de Jean Dumas, aussi somptueuses nous apparaissent-elles plus de trente ans après, aient été, a priori, suspectes dans le milieu où il évoluait. Quant à sa démarche analytique, pourtant si clairement contestée par W. Lemit, elle avait des antécédents prestigieux (Etay dans Rocher, 1994 : Annexe 1, p.159).

Après avoir préalablement vanté les mérites de la recherche d'authenticité de notre collecteur, il faut tout de même noter que ce corpus présente encore des défauts inhérents à ses conditions écologiques d'enregistrement. On retrouve par exemple des éléments perturbateurs qui viennent interrompre le déroulement des chansons ou bruitent les audios. On rencontre au fil des écoutes des cris d'animaux : poules, coqs (JD_a330_02, 44), canaris, chiens (JD_a330_04, 23 & 25) ; des bruits de bouche extra-linguistiques : raclement de gorge, crachat, toux (JD_a064_12, 45-46) ; des bruits domestiques : carillon (JD_a330_02, 48-49), clocher, klaxons (JD_a156_02, 21), eau qui coule ; des bruits liés aux activités humaines : fuseaux qui s'entrechoquent (dentelle au fuseau ; JD_a354_05), enfants (JD_a328_05, 8), bavardages... Des problèmes techniques viennent de surcroît s'ajouter à cela : traces de repiquage, effacement d'une partie de certaines chansons... Jean Dumas fait aussi face à des chanteurs qui oublient leurs paroles, brisant le bon déroulement de la chanson, ou à d'autres qui refusent de chanter la suite de chansons grivoises[12], privant ainsi l'auditeur d'une partie de la chanson. D'autres biais potentiels sont l'enregistrement avec des chanteurs alcoolisés, le travail avec des témoins qui semblent majoritairement bilingues, un biais de l'écrit qui n'est pas à sous-estimer (une enquêtée dit qu'elle copie ses chansons à sa sœur, une autre chante et on entend des bruits de pages tournées).

Pour terminer, il me fallait des documents sonores écologiques, comprenant du chant, en français, contextualisables, ce que proposait le FJD. Je ne saurais expliquer pourquoi je me suis principalement concentrée sur ce fonds plutôt qu'un autre. Peut-être était-ce le premier à remplir tous les critères susmentionnés, malgré quelques défauts que nous venons d'aborder, et que ce fonds était déjà relativement important dans le cadre d'une thèse. Finalement, il s'est avéré qu'il avait été enregistré dans des régions qui m'étaient familières.

Ainsi, la démarche de Jean Dumas présente des défauts : il économise la surface écrite autant que les bobines, coupe drastiquement les passages non chantés, ne collecte que très peu de chants en langue régionale, manque parfois d'éthique des relations interpersonnelles lors des séances, renforçant le paradoxe de l'observateur, et agit en grande partie en solitaire dans le monde des collecteurs de chants de tradition orale. Les enregistrements qu'il produit sont enfin bruités et peuvent être biaisés par divers facteurs.

[12] Selon un processus d'**édulcoration**, ou « [p]rocessus qui consiste à sectionner un « fragment » du texte d'origine afin d'atténuer les contenus jugés grivois » (Surmont, 2010 : 153).

En guise de conclusion, j'ai découvert avec enthousiasme la richesse des fonds ethnomusicologiques déjà numérisés. Ma quête d'un corpus de chansons de tradition orale authentiques m'a poussée à travailler sur le fonds Jean Dumas, disponible sur la Base inter-régionale du patrimoine oral, collecté dans les années 1960 dans le centre de la France par Jean Dumas, un professeur d'italien amateur de musique, et son épouse. La fine sélection des chansons par notre collecteur et sa décision d'utiliser l'enregistrement sonore font qu'on en oublierait les quelques défauts inhérents à sa démarche de recueil.

Le corpus suivant dont il sera question, un corpus de chansons issues de l'industrie du disque, nécessite une transition, afin de comprendre en quoi ce nouvel objet diffère de chansons de tradition orale. Avec les chansons issues de l'industrie du disque ou chansons médiatisées (Zumthor, 1986), l'objet chanson est redéfini, comme avec la *règle des trois minutes* (« aucune chanson ne doit dépasser cette durée »), imposée par l'industrie du disque (Zumthor, 1986 : 14), qui renégocie drastiquement la durée des chansons, auparavant très variable dans la tradition orale. Par ailleurs, l'expérience de l'auditeur et son lien avec le chanteur sont tout à fait bouleversés : d'une co-présence spatio-temporelle, on passe à une séparation spatio-temporelle des différents protagonistes (Koch & Oesterreicher, 2001 : 586), et la co-construction de l'œuvre artistique par le chanteur et ses auditeurs n'a plus cours (Zumthor, 1986 : 17). Le son enregistré favorise une consommation individuelle, auparavant collective, augmente considérablement le nombre potentiels d'auditeurs et leur laisse plus de liberté, ceux-ci pouvant choisir quelle œuvre écouter, quand, où et dans quelles circonstances pratiquer cette écoute (Zumthor, 1986 : 17). Les chansons de l'industrie du disque sont enfin des **chansons signées**, ou « [o]bjet[s]-chanson[s] dont l'identité de l'auteur et du compositeur de musique est connue et/ou identifiable », contrairement aux chansons de tradition orale (Surmont, 2010 : 153). Après ces précisions, faisons pour l'heure connaissance avec un autre corpus, dit de tops 1.

3.3 Corpus de tops 1

Afin de faire une étude sur corpus du FJD, je devais créer des données secondaires, ce qui était nouveau pour moi puisque je m'étais arrêtée à l'étape de collection d'exemples pour mes recherches de master comme je l'ai mentionné en 2.1.1. J'avais aussi été brièvement initiée au logiciel CLAN en master, mais je souhaitais perfectionner son maniement en transcrivant et annotant un premier corpus. À la fin de l'année universitaire 2017-2018, soit en fin de première année de thèse, j'ai eu l'idée de constituer un corpus

de tops 1, et je me suis dit que cela pouvait être une occasion d'entraînement avant de commencer à traiter le FJD. L'intérêt scientifique à rassembler de telles chansons en corpus était de proposer de nouvelles données microdiachroniques pour l'étude de la liaison, car comme on l'a vu en 1.2.3.1.1.4, dans la parole publique, seuls des personnalités politiques et des journalistes avaient jusqu'alors fait l'objet de recherches. En synchronie, je n'avais connaissance que d'un unique corpus chanté mobilisé pour l'étude de la liaison, constitué d'enfantines et que j'ai mentionné en 1.2.3.4.3. De surcroît, il ne me semblait pas que ce nouveau corpus ait déjà été constitué à d'autres fins de recherche en linguistique.

Ce corpus de chansons ayant accédé à la première place de classements musicaux réalisés pour la France entre 1956 et 2017, d'une durée de 21h 38min 49s, comporte ici 368 chansons (et non plus 366 comme lors d'une précédente étude (Coutanson & Badin, 2021 : 136), puisque j'ai réintégré deux chansons qui causaient de multiples annotations erronées), et comptabilise 112 123 mots[13] pour 5 973 contextes de liaison possibles. Les chansons amassées « [m'ont semblé pertinentes] en tant qu'input linguistique pour les auditeurs français et [représentatives] du champ musical d'une période donnée » (Coutanson & Badin, 2021 : 136). Je n'ai pas essayé de constituer un corpus socialement équilibré de la communauté des chanteurs d'expression française du champ musical français, car selon mon critère de sélection, obtenir tous les titres numéro un sur une période donnée, l'ensemble retenu est exhaustif. Je peux néanmoins facilement obtenir des informations sociolinguistiques concernant les chanteurs de ce corpus.

En ce qui concerne les chanteurs pour commencer, certains titres sont interprétés en solo, d'autres en duo, quand d'autres encore sont l'œuvre de collectifs présentant de multiples chanteurs. La figure 3.5 présente pour chaque chanteur le nombre de titres que ce dernier a réussi à faire accéder à la première place des classements. Attention, leur nombre est trop important pour que tous ces protagonistes puissent être visibles en abscisse. C'est Johnny Hallyday, avec vingt-quatre chansons s'étant hissées en top 1, Mylène Farmer (avec 15 chansons), Michel Sardou (avec 13 chansons), Dalida (avec 12 chansons), Claude François (avec 8 chansons), Salvatore Adamo et Sheila (avec 7 chansons chacun), Mike Brant et Richard Anthony (avec 6 chansons chacun) et Céline Dion et Charles Aznavour (avec 5 chansons chacun) qui tiennent la tête du classement. Un petit nombre de chanteurs est donc habitué aux succès musicaux, quand « un très grand nombre de chanteurs n'atteindra qu'une fois la première place du hit-parade (43% ; 173/402) » (Coutanson & Badin, 2021 : 137). Dans la figure 3.6, on peut noter

[13] Voir la note de bas de page 7. Ici, j'ai additionné le nombre de tokens pour SGR (108471), SGR2 (3615) et SGR3 (37).

plus d'hommes (67% ; 270/403) que de femmes (33% ; 133/403) en tête des classements (Coutanson & Badin, 2021 : 137). L'âge de ces chanteurs est présenté dans la figure 3.7, sur une échelle allant de 0 à 99 ans (tous les âges en abscisse ne peuvent être affichés), « avec une fourchette allant de 4 (Jordy avec *Dur dur d'être bébé !*, 1992 [1992]) à 65 ans (Johnny Hallyday avec *Ça n'finira jamais*, 2008 [2008]) » (Coutanson & Badin, 2021 : 137). La majorité légale est nettement marquée sur le graphique, même si on trouve certains mineurs. La plupart des interprètes avaient entre 20 et 35 ans lorsque leur chanson a été élevée à la première place des classements. Enfin, ceux-ci ne sont pas tous nés en France (figure 3.8a[14]) et certains ne sont pas de langue maternelle française. Ceux qui sont nés en France proviennent pour la plupart de la région parisienne (figure 3.8b[14]), grande pourvoyeuse de talents musicaux.

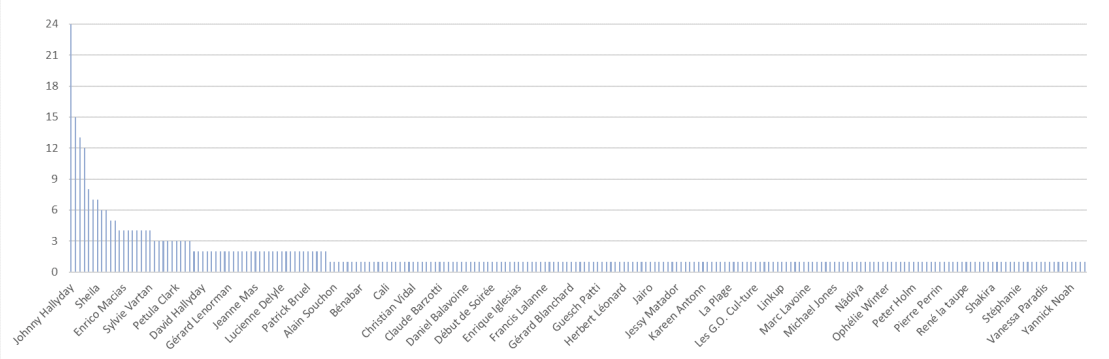


FIGURE 3.5 : Nombre de chansons en top 1 par chanteur

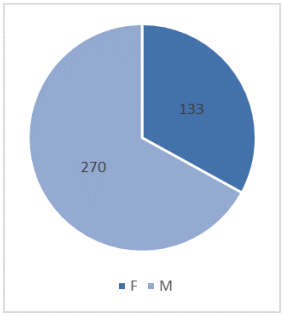


FIGURE 3.6 : Sexe des chanteurs du corpus de tops 1

En ce qui concerne les titres ensuite, on voit dans la figure 3.9 qu'entre un (en 2012) et treize (en 2007) titres francophones ou partiellement francophones se sont

[14]Les deux cartes ont été réalisées à l'aide du logiciel de cartographie propriétaire BatchGeo, accessible en ligne à l'adresse suivante : https://fr.batchgeo.com/.

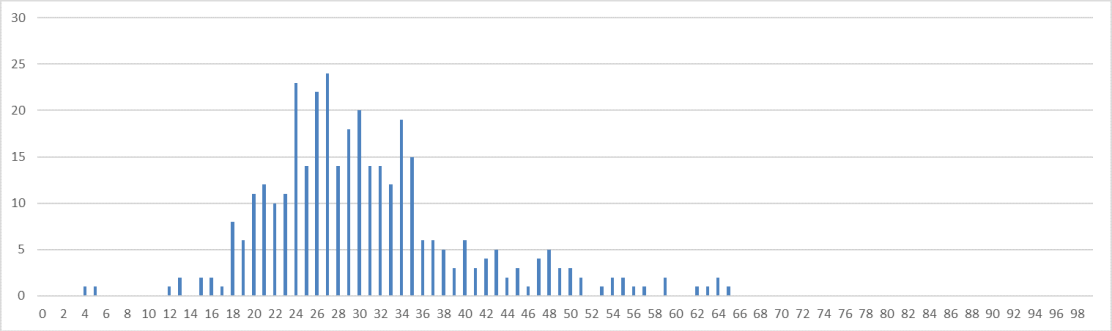


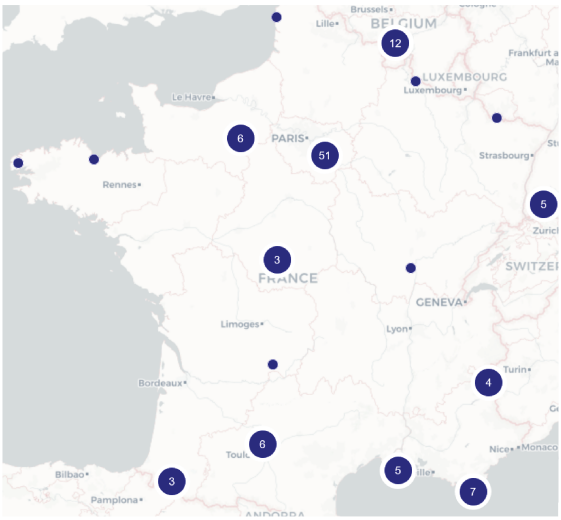
FIGURE 3.7 : Âge des chanteurs du corpus de tops 1

hissés à la première place des classements (Coutanson & Badin, 2021 : 137). Cette variation peut être due à deux facteurs : un titre francophone (voire plusieurs) s'est maintenu longtemps cette année-là en tête des classements et/ou des titres interprétées dans d'autres langues ont fait barrage cette même année aux titres francophones. Dans la figure 3.10, on peut noter qu'une minorité de chansons s'est maintenue dans la durée en tête des classements, alors qu'une majorité des titres n'y a fait qu'un bref passage, n'excédant souvent pas la semaine (Coutanson & Badin, 2021 : 137).

Dans cette partie, je suis tout d'abord revenue sur l'intérêt de constituer ce corpus de tops 1. Je l'ai par la suite présenté, en insistant sur son caractère clos, ce qui m'a évité de devoir recourir à un échantillonnage représentatif afin de constituer un corpus équilibré. J'ai pour finir présenté le profil sociolinguistique des chanteurs qui composaient ce corpus, avant de commenter quelques caractéristiques des titres concernés.

Dans ce chapitre, j'ai tout d'abord mentionné l'existence des Archives de la Parole, ces premières archives sonores dont on dispose pour le français. J'ai pour continuer amplement présenté le fonds Jean Dumas, avant de résumer comment les chansons de l'industrie du disque avaient révolutionné l'objet chanson. Ultérieurement, j'ai présenté un nouveau corpus que j'ai constitué dans le cadre de cette thèse, nommé corpus de tops 1. Le prochain chapitre se consacrera aux résultats de ce travail de thèse et à une discussion, en commençant par le carottage.

(a) Vue clustérisée mondiale



(b) Vue clustérisée hexagonale

FIGURE 3.8 : Lieu de naissance des chanteurs du corpus de tops 1

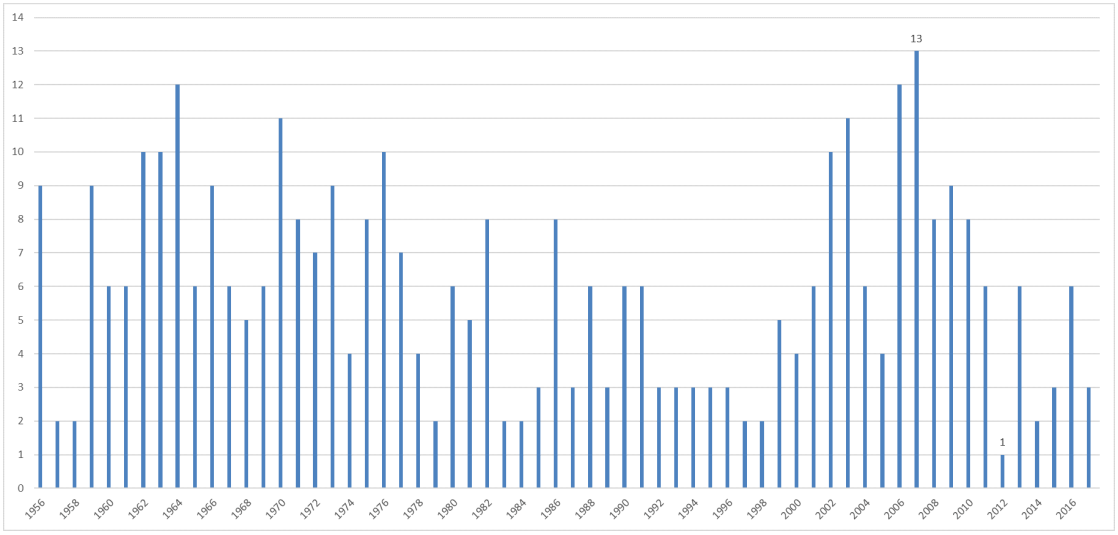


FIGURE 3.9 : Nombre de tops 1 (partiellement) francophones par année

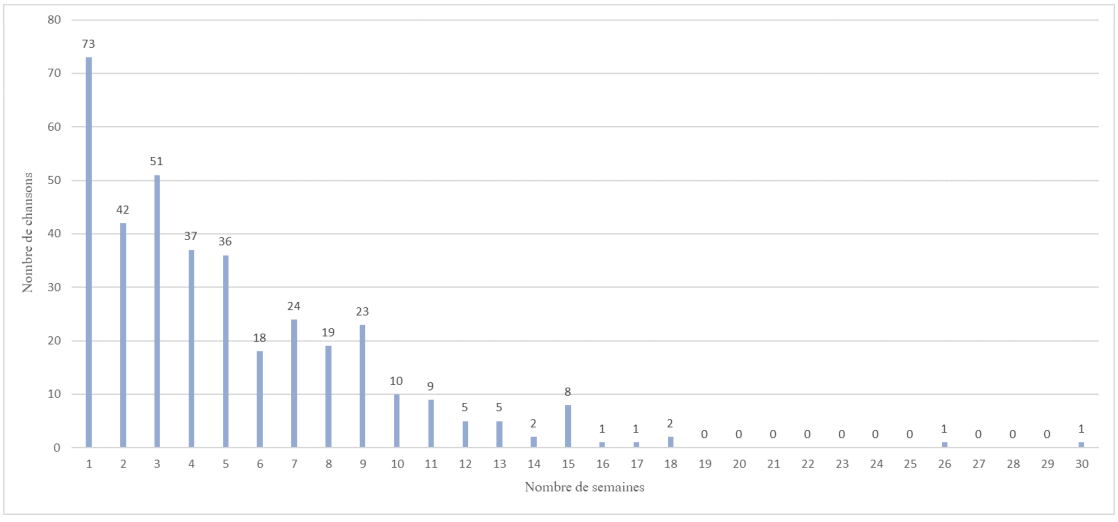


FIGURE 3.10 : Nombre de chansons selon la durée de maintien au top 1

Chapitre 4

Résultats et discussion : carottage et première fouille des corpus

4.1 Carottage des corpus choisis

Comme indiqué en 2.1.2, j'ai sélectionné cinq corpus à carotter : le FJD, PFC, ALIPE, l'ALAVAL et le fonds Charles Joisten. En plus du FJD sur lequel était fondée ma collection d'exemples de pataquès dans mes mémoires de recherche de master (2.1.1), j'ai retenu les corpus PFC et ALIPE, deux corpus annotés pour l'étude de la liaison[1]. Je souhaitais également un corpus en francoprovençal et j'ai finalement opté pour l'ALAVAL, un atlas linguistique audiovisuel dit semi-spontané (Kristol, 1998) et j'ai rajouté à cela le fonds Charles Joisten, avec pour objectif d'avoir un deuxième corpus ethnomusicologique pour éviter de généraliser mes observations sur le FJD à tous les corpus ethnomusicologiques. Au départ, je comptais également carotter les deux volets du corpus des ESLO, un grand corpus variationniste du français (voir Eshkol-Taravella et al., 2011 pour une présentation du corpus), et les deux volets du corpus HPOL, un corpus de parole publique provenant de discours de personnalités politiques de 1908 à 2015, annoté pour la liaison (Laks & Peuvergne, 2017), mais j'ai dû y renoncer, faute de temps. De plus, une première fouille de ces deux derniers corpus n'a pas été possible car les annotations d'ESLO pour la liaison ne sont ni systématiques ni librement accessibles et les données primaires d'HPOL n'étaient pas accessibles en 2018.

[1] Comme le note Blanche-Benveniste, 1997 : 30, c'est surtout les liaisons ou absences de liaisons remarquables qui pourraient intéresser les linguistes. Si bon nombre de corpus oraux ne présentent pas d'annotation systématique pour la liaison, certaines liaisons atypiques pourraient y avoir été annotées, comme dans CFPP2000.

J'ai donc retenu le FJD, que je connaissais bien, deux corpus annotés pour la liaison, PFC et ALIPE, un atlas linguistique en francoprovençal, l'ALAVAL et un deuxième corpus ethnomusicologique, le fonds Charles Joisten, en laissant finalement de côté les corpus ESLO et HPOL en raison de leur état d'avancement. Ces corpus sont présentés ci-dessous, avec comme premier candidat le fonds Jean Dumas.

4.1.1 Présentation des corpus carottés

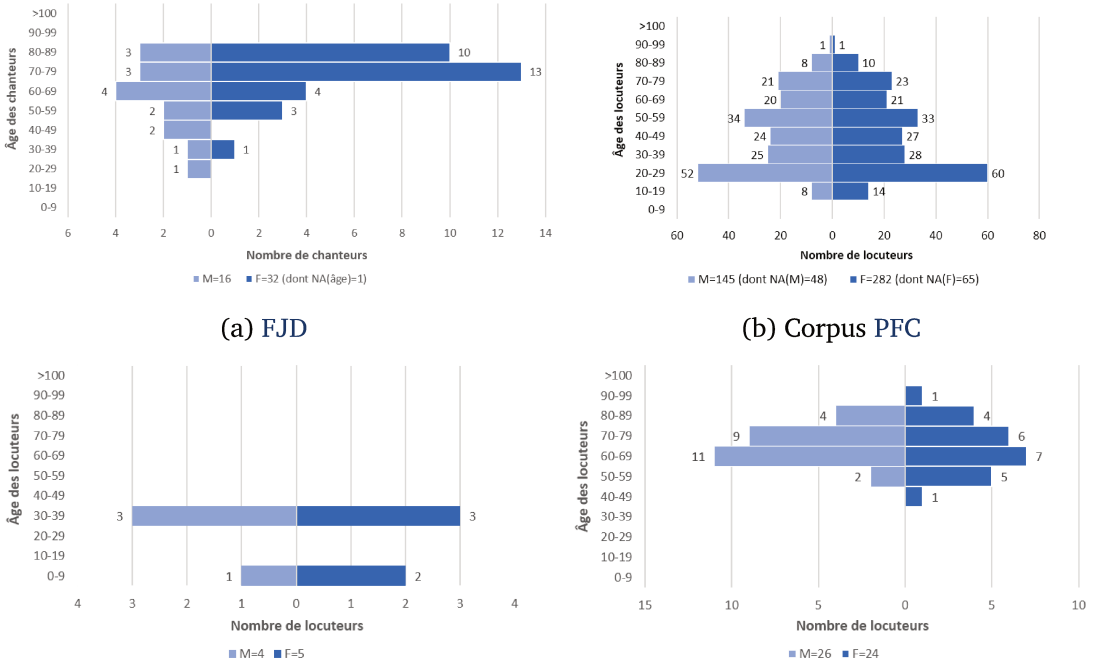
Les corpus sont rapidement présentés dans le tableau 4.1. Afin de mieux appréhender les différentes populations d'étude, les pyramides des âges des locuteurs des corpus sont mises en avant dans la figure 4.1. Enfin, le tableau 4.2 nous informe sur ce que je nomme l'état d'avancement des corpus, qui permet de se faire une idée de l'état dans lequel se trouvent les corpus qu'on souhaite utiliser selon l'étude à effectuer.

Nom du corpus	**FJD**	**PFC**	**ALIPE**	**ALAVAL**	**Fonds Charles Joisten (chansons uniquement)**
Type de corpus	Ethnomusicologique	Variationniste	Acquisitionnel, interactionnel	Atlas linguistique audiovisuel	Ethnomusicologique
Langue	français (principalement), occitan	français	français	francoprovençal	français, francoprovençal, occitan
Date d'enregistrement	1958-1962	1996-2015	2014 ?	1994-2001	1957
Lieu d'enregistrement	Haute-Loire, Puy-de-Dôme	Pays francophones	France hexagonale (domicile familial)	Valais (Suisse)	Hautes-Alpes
Nombre de locuteurs	47	435	3 enfants + 6 parents	50	NA
Âge des locuteurs	de 24 à 83 ans	de 10 à 93 ans	de 2;4 à 5;4 ans	de 47 à 92 ans	NA

TABLEAU 4.1 : Informations relatives aux corpus carottés

4.1.1.1 Fonds Jean Dumas

Il peut sembler étonnant de carotter le FJD, un fonds que je connais bien depuis mes recherches de master. Pourtant, je voulais voir ce qui pouvait ressortir de chaque corpus avec une méthodologie commune. Comme précisé en 3.2.2, je n'ai travaillé que sur les quatre premiers corpus du FJD qui étaient disponibles à cette époque (voir tableau



(a) FJD

(b) Corpus PFC

(c) Corpus ALIPE

(d) ALAVAL

FIGURE 4.1 : Pyramides des âges des différents corpus carottés

3.1). Les informations relatives à ce fonds avaient pour la majorité déjà été présentées en 3.2. Rappelons seulement que les données secondaires, principalement des PDF de chansons manuscrites, ne sont pas très pratiques et n'ont pas été utilisés, que seules certaines non-réalisations de liaisons ou réalisations de pataquès y étaient indiquées et que les données n'ont pas été anonymisées (voir tableau 4.2).

Dans le tableau 4.1 et la figure 4.1, les informations relatives au nombre, au sexe et à l'âge des locuteurs de ce fonds sont bien présentes. Anticipant une étude sur corpus du FJD, je suis allée rechercher ces informations sur les pages de chaque corpus du fonds (BIPO, 2022b,c,g,h) pour les deux premières informations, ainsi que sur une fiche manuscrite rédigée par Jean Dumas pour chaque chanteur afin d'obtenir leur âge. Les chanteurs du FJD étaient assez âgés, et de nombreuses femmes y ont été enregistrées, comme on peut le voir dans la figure 4.1a.

4.1.1.2 Corpus Phonologie du Français Contemporain

Le corpus variationniste PFC, recueilli dans les divers pays francophones entre la fin des années 1990 et les années 2010, présente le désavantage de ne pas être entièrement

Nom du corpus	FJD	PFC	ALIPE	ALAVAL	Fonds Charles Joisten (chansons uniquement)
Accès libre aux données primaires ?	Oui	En partie seulement	Oui	Oui	Oui
Classification	Par corpus, par chanteur, par enquête, par chanson	Par enquête, par locuteur, par type de discours...	Par interaction	Morphosyntaxique	Par chanteur, par enquête, par chanson
Transcrit ?	Oui (parfois absente)	Oui (10min par locuteur en conversation guidée + 10min en conversation libre)	Oui	Oui	Oui (parfois absente)
Type de transcription	Orthographique : manuscrite, tapuscrite (corpus 3)	Orthographique : tapuscrite	Orthographique : tapuscrite	Phonétique : tapuscrite	Orthographique : manuscrite
Format de transcription	.pdf	.TextGrid (Praat)	.cha (CLAN)	.pdf	.pdf
Anonymisé ?	Non	En partie seulement	Oui	Non	Non
État de l'annotation	Certains pataquès sont annotés	Annoté pour la liaison (5min + 5min) + certains pataquès sont annotés	Annoté pour la liaison + certains pataquès sont annotés	Liaisons et pataquès présents dans la transcription	Non annoté
Accès libre aux données secondaires ?	NA	Oui	Oui	Non	NA
Outil d'exploration liaison/ pataquès ?	Non	Oui	Oui	Non	Non

TABLEAU 4.2 : État d'avancement des corpus carottés

libre d'accès, car des transcriptions n'ont pas été anonymisées. Celles qui l'ont été sont disponibles sur ORTOLANG[2] (Modèles, Dynamiques, Corpus - UMR 7114 (MoDyCo) & Université de Groningen (RUG), 2017), sous réserve de l'acceptation de la charte ORTOLANG et de la licence relative à ce corpus. Cela ne m'a pas posé de difficulté car je disposais des accès à la base de données PFC[3], réservés aux chercheurs en ayant fait la demande, qui comprend à la fois les données anonymisées et les données non anonymisées.

Ce corpus n'est pas transcrit et annoté dans sa totalité. En effet, comme indiqué dans

[2]Elles sont accessibles en suivant l'identifiant pérenne suivant : https ://hdl.handle.net/11403/pfc/v1.

[3]La base PFC est accessible à l'adresse suivante : https ://research.projet-pfc.net/.

le tableau 4.2 et selon la méthodologie mise en place lors de la constitution du corpus PFC, seules dix minutes de conversation guidée et dix minutes de conversation libre devaient être transcrites, et seules cinq minutes pour chacune devaient être codées pour la liaison, quand il était conseillé d'enregistrer des entretiens guidés d'une vingtaine de minutes et des entretiens libres d'une trentaine de minutes (Durand et al., 2002). Ainsi, la richesse du corpus n'est pas pleinement exploitée (rappelons toutefois que les linguistes se méfient des premières minutes de captation, le temps que les locuteurs se sentent à l'aise avec le dispositif d'enregistrement, et que ces consignes ont également permis d'obtenir des extraits temporels comparables). Les données primaires ne peuvent ainsi pas être intégralement fouillées à l'aide des données secondaires.

Dans le tableau 4.1 et la figure 4.1, les informations relatives au nombre, au sexe et à l'âge des locuteurs sont bien renseignées. Celles-ci ont été récupérées depuis la plateforme de présentation du corpus, mais il est important de préciser que les informations disponibles en juin 2018, lors de ma phase de carottage des corpus, étaient lacunaires, notamment au niveau de l'âge des locuteurs, qui n'était pas systématiquement indiqué (voir figure 4.1b). La pyramide des âges de ce corpus est plus équilibrée (voir figure 4.1b), même si on trouve beaucoup de jeunes dans la vingtaine, ce qui pourrait venir du fait que les enquêteurs sont souvent des étudiants qui pourraient se tourner préférablement vers des locuteurs de leur âge, puis vers les générations de leurs parents voire grands-parents selon la méthode boule de neige. La très grande majorité des locuteurs de ce corpus étaient majeurs, puisqu'il ne s'agit pas d'un corpus acquisitionnel. Seuls deux locuteurs ont moins de 16 ans : il s'agit de 54bmd1 (10 ans) et caalg1 (14 ans). PFC étant un corpus variationniste, il était prévu dans la méthodologie initiale (Durand et al., 2002) qu'il soit équilibré hommes/femmes, ce qui n'est pas tout à fait le cas parmi les participants au projet (voir figure 4.1b), bien que la différence ne soit pas significative.

4.1.1.3 Corpus Acquisition de la Liaison et Interaction Parents Enfants

Venons-en au corpus ALIPE, un corpus acquisitionnel, interactionnel, dense et recueilli en situation écologique par les parents dans les années 2010 (Liégeois, 2014). Les données primaires et secondaires d'ALIPE sont librement accessibles depuis ORTOLANG[4] (Chabanal et al., 2017), sous réserve de l'acceptation de la charte ORTOLANG et de la licence relative à ce corpus. Ce dernier est entièrement transcrit et annoté pour la

[4]Elles sont accessibles en suivant l'identifiant pérenne suivant :
https ://www.ortolang.fr/market/corpora/alipe-000853/v1.1.

liaison, et l'audio et les transcriptions ont été anonymisées, comme le mentionne le tableau 4.2.

Pour ce corpus, les informations relatives au nombre, à l'âge et au sexe des locuteurs sont issues de la thèse de Loïc Liégeois (2014 : 173, 175). La figure 4.1c montre que les locuteurs de ce corpus appartiennent à deux générations : celle des parents, qui ont tous la trentaine, et celle des enfants. L'âge précis de ces derniers à chaque période d'enregistrement est indiqué dans le tableau 4.3 (selon le format années;mois.jours). Les enfants étant issus de couples hétérosexuels, il y a autant de mères que de pères, mais seulement un garçon pour deux filles. Enfin, notons au passage que les parents sont issus des classes moyenne ou moyenne-supérieure en terme d'Indice de position socioéconomique (IPSE) (Liégeois, 2014 : 175).

Période d'enregistrement	Salomé	Baptiste	Prune
T1	2;4.04 - 2;4.08	2;11.18 - 2;11.25	3;4.07 - 3;4.13
T2	2;11.28 - 3;0.03	3;5.27 - 3;6.11	4;0.01 - 4;0.03
T3	NA	NA	5;4.10 - 5;4.15

TABLEAU 4.3 : Âge des enfants dans le corpus ALIPE à chaque période d'enregistrement (Liégeois, 2014 : 173)

4.1.1.4 Atlas Linguistique du Francoprovençal Valaisan

Les données primaires et secondaires de l'ALAVAL, que les responsables du projet ont décidé de ne pas anonymiser, proviennent pour leur part du site de présentation des données[5]. Les données de cet atlas, recueilli dans le canton du Valais, en Suisse, sont présentées principalement selon une classification morphosyntaxique, mais aucun accès aux données primaires brutes n'est proposé, c'est-à-dire aux enquêtes dialectologiques, hors extraits. En 2018, une correspondance avec Andres Kristol, un des initiateurs de cet atlas, m'a permis d'apprendre que ces documents bruts, vieillissants, n'étaient toujours pas numérisés pour l'heure, et que l'atlas continuait à être alimenté. En outre, comme le présente le tableau 4.2, l'équipe de chercheurs ayant travaillé sur l'atlas a opté pour une transcription phonétique (segmentée en mots) et non orthographique, ce qui se prêtait bien à la variabilité du francoprovençal, mais empêche de rechercher un lexème en particulier, à moins de passer par la traduction en français, ce qui n'est pas toujours adéquat. Cela aura tout de même permis une transcription systématique des liaisons et pataquès, souvent séparément du M1 et du M2. Enfin, le site de présentation

[5]Le site du projet est accessible à l'adresse suivante : http ://alaval.unine.ch/.

des données mis en place ne permet malheureusement pas d'exporter toutes les données dans un tableur, mais seulement les données pour un contexte morphosyntaxique spécifique, au format PDF.

Concernant les informations quant au nombre, au sexe et à l'âge des locuteurs de cet atlas linguistique, elles proviennent de l'introduction au projet (Berchtold et al., 2019b : 29-45). Dans le cadre du travail sur une Langue en danger (LED), les locuteurs enregistrés sont pour la plupart âgés (la plus jeune locutrice avait 47 ans en 1997), comme l'illustre la figure 4.1d. L'ALAVAL se voulait équilibré hommes/femmes, ce qui a presque été réalisé, comme l'indiquent les enquêteurs : « [d]ans chaque point d'enquête, nous avons enregistré deux témoins (une femme et un homme, à l'exception de Vouvry où nous n'avons plus trouvé de locutrice du parler local et avons donc enregistré deux hommes) » (Berchtold et al., 2019b : 20).

4.1.1.5 Fonds Charles Joisten

Enfin, le fonds Charles Joisten (BIPO, 2022e), un fonds en partie ethnomusicologique provenant également de la BIPO (comme mentionné en 2.1.2), recueilli par un collecteur nommé Charles Joisten, a été enregistré à la même époque que le FJD, mais malheureusement dans une région assez proche de la région de recueil du premier fonds, bien que l'aire en question soit plus francoprovençale qu'occitane cette fois-ci. Sur la BIPO, j'ai à nouveau tenté d'avoir accès aux données primaires de ce fonds (les documents sonores), ainsi qu'à de premières données secondaires (paroles de chansons et partitions manuscrites), qui n'ont pas été anonymisées. Ces dernières n'étaient pas systématiquement présentes, et ni les liaisons ni les pataquès n'y étaient annotés. Je souhaitais également avoir accès à des informations sociologiques concernant les chanteurs, mais elles se sont révélées bien maigres pour ce fonds : seulement le prénom et le nom de chaque chanteur (on peut à nouveau en déduire son sexe), ainsi que le lieu d'enregistrement (pouvant potentiellement correspondre au lieu de vie du chanteur).

Dans le tableau 4.1 et la figure 4.1, le nombre de locuteurs, leur âge et leur sexe font défaut. Comme je le présenterai dans le tableau 4.2, j'ai décidé de ne pas analyser plus finement ce fonds, et contrairement au FJD, je n'ai pas pris la peine de recenser ces informations sans compter que, comme je viens de le mentionner, l'âge des locuteurs faisait défaut.

En résumé, lors de cette phase de carottage, les données primaires du FJD et du fonds Charles Joisten, non anonymisées, ont été obtenues depuis la BIPO. Pour PFC,

je me suis servie de données primaires et secondaires n'étant pas entièrement anonymisées, issues d'une base de données réservée aux chercheurs. Il faut se rappeler que les données primaires de PFC ne sont pas intégralement transcrites et annotées pour la liaison. Pour ALIPE, les données primaires et secondaires anonymisées proviennent d'ORTOLANG. Pour finir, les données primaires et secondaires de l'ALAVAL, non anonymisées, proviennent de l'atlas en ligne. Cependant, les enquêtes dialectologiques complètes ne sont pas accessibles et la transcription phonétique tout comme l'export par contexte morphosyntaxique et en .pdf ne facilitent pas la fouille de l'atlas. Pour les cinq corpus sélectionnés, je suis allée rechercher les informations relatives au nombre, à l'âge et au sexe des locuteurs, à l'exception du fonds Charles Joisten. Les chanteurs du FJD sont assez âgés. Pour PFC, toutes les classes d'âge ne comprennent pas autant de locuteurs, même si la pyramide des âges reste relativement équilibrée. En ce qui concerne ALIPE, les enfants ont été enregistrés entre leurs deux et cinq ans. Leurs parents, trentenaires, appartenaient aux classes moyenne ou moyenne-supérieure. Enfin, le francoprovençal étant une LED, l'ALAVAL a été recueilli chez des locuteurs majoritairement âgés, mais l'atlas est bien équilibré hommes/femmes. Maintenant que nous avons fait connaissance avec les corpus à carotter, place à la méthodologie de carottage, voire de première fouille des données que je présenterai dans la sous-partie suivante.

4.1.2 Méthodologie de carottage, voire de première fouille des données

Dans un premier temps, je cherchais à effectuer une première fouille des données pour les corpus déjà annotés pour la liaison, soit PFC et ALIPE, en recherchant des contextes préférentiels d'apparition de pataquès, tout en contrôlant le bruit (ici, les occurrences annotées à tort comme pataquès), afin de constituer une collection d'exemples par corpus.
PFC peut être fouillé de deux façons (Eychenne et al., 2016 : 390, 392, 398) :

- directement sur l'interface de présentation du corpus (https ://research.projet-pfc.net/), à l'aide du moteur de recherche associé qui liste les occurrences issues de la requête, en ligne, et permet leur téléchargement dans un tableur
- en utilisant le programme Dolmen, qui propose d'effectuer des requêtes à l'aide d'une interface utilisateur simplifiée ou à l'aide d'expressions régulières, les résultats s'affichant à nouveau directement dans l'interface ou pouvant s'exporter dans un tableur.

J'ai opté pour la première solution, qui était plus rapide (nul besoin d'obtenir les données au préalable) et suffisante à ce niveau.

Pour ALIPE, j'ai communiqué avec Loïc Liégeois lors de cette étape (en 2018), qui m'a réexpliqué le codage de la liaison utilisé dans sa thèse, tout en me conseillant des requêtes pour fouiller le corpus à l'aide de CLAN.

Dans un second temps, je souhaitais prélever plusieurs carottes dans les corpus, d'une durée totale de vingt minutes. Je ne souhaitais pas me limiter aux pataquès, mais également observer le comportement des liaisons dans chaque corpus ou d'autres phénomènes de sandhis externes comme les harmonies consonantiques en frontière de mots. Après une première écoute des extraits me permettant d'obtenir une collection d'exemples des phénomènes de sandhis observés pour chaque corpus, je souhaitais faire converger les données secondaires de ces corpus vers des transcriptions CLAN codées pour la liaison et le pataquès comme mentionné dans la partie méthodologique (2.1.3.4).

Pour PFC, j'avais décidé de choisir quatre extraits de cinq minutes par locuteur, issus de conversations libres ou guidées. Je recherchais deux hommes et deux femmes, avec un locuteur résidant en France hexagonale, un locuteur au Canada, un locuteur en Afrique subsaharienne et un locuteur en Suisse ou en Belgique. Il fallait aussi que les locuteurs appartiennent à des classes d'âges distinctes : le premier devait avoir entre 20 et 29 ans, le second entre 30 et 49 ans, le troisième entre 50 et 69 ans et le dernier plus de 70 ans. Pour obtenir des transcriptions/annotations similaires, cela impliquait pour PFC de convertir les transcriptions de Praat vers CLAN et de réannoter les liaisons.

Pour ALIPE, j'envisageais de consacrer sept minutes à chaque enfant (ce corpus comprend trois enfants), prises au hasard dans le corpus, voire plus car j'avais prévu de déduire les interventions des parents. ALIPE posait moins de difficultés pour l'uniformisation des transcriptions/annotations, car ce corpus a été transcrit et annoté sous CLAN. Seul un recodage des liaisons et pataquès était nécessaire.

En ce qui concerne l'ALAVAL, la plateforme de présentation de cet atlas linguistique propose un classement par contexte morphosyntaxique. Pour chaque contexte, je comptais me focaliser sur l'énoncé de même rang que le contexte en question (pour le premier contexte, le premier énoncé proposé ; pour le second contexte, le second énoncé…), jusqu'à atteindre une durée totale de vingt minutes, en contrôlant les locuteurs afin qu'ils soient assez diversifiés. Dans ma quête de transcriptions/annotations uniformisées à la fin de cette étape de carottage, j'ai fait une exception pour l'ALAVAL. En effet, la mise en avant d'énoncés plutôt que des enquêtes complètes, leur transcription phonétique et le format de l'export en PDF m'ont dissuadée dès le départ de transcrire et annoter ces productions sous CLAN, sachant également dès lors que je n'allais pas dépasser l'étape de carottage pour cet atlas.

Pour le FJD et le fonds Charles Joisten, je prévoyais enfin de sélectionner des chansons au hasard à carotter, tout en vérifiant que celles-ci soient de chanteurs différents, jusqu'à parvenir à une durée totale de vingt minutes. Le FJD et le fonds Charles Joisten n'étant pas transcrits, les chansons sélectionnées dans ces corpus devaient pour leur part être intégralement transcrites et annotées sous CLAN.

Pour résumer, la méthodologie mise en place lors de cette étape se subdivise en deux temps : un premier pour fouiller les corpus déjà annotés pour la liaison et le pataquès, à l'aide de requêtes sur le site de PFC et de requêtes CLAN pour ALIPE, un second pour carotter vingt minutes des cinq corpus sélectionnés, en prélevant diverses carottes au hasard, tout en contrôlant parfois certaines caractéristiques sociologiques des locuteurs et que ces derniers soient assez diversifiés, tout en uniformisant les transcriptions/annotations (conversion des transcriptions au format CLAN et réannotation des liaisons). Il est temps de découvrir ensemble ce qui peut ressortir de cette phase de carottage et de première fouille des données.

4.1.3 Résultats du carottage

Le tableau 4.4 présente le nombre de pataquès relevés lors du carottage de chaque corpus, par ordre croissant. Il faut être très prudent, et ne pas prendre ces données pour ce qu'elles ne sont pas. Un carottage étant un échantillonnage non représentatif, le seul mérite de ce relevé est de nous apprendre que tous les corpus pourraient présenter des pataquès (en ce qui concerne ALIPE, outre une phase de carottage, le format de transcription et d'annotation du corpus permet une première fouille des données qui, comme nous le verrons, s'est révélée fructueuse), mais il est trop tôt pour savoir dans quelle mesure chacun d'eux en contient. Nous détaillerons à présent ces résultats pour chaque corpus, en commençant par l'ALAVAL.

	ALIPE	**PFC**	**Fonds Charles Joisten**	**FJD**	**ALAVAL**
Durée de carottage	20 min	20 min	20 min	20 min	10 min
Nombre d'occurrences	0	2	2	5	4

TABLEAU 4.4 : Nombre de pataquès par carotte (modifié depuis Coutanson, 2020 : 9 après réanalyse)

4.1.3.1 Résultats du carottage de l'ALAVAL

Comme on le voit dans le tableau 4.4, si la durée de carottage de l'ALAVAL avait été fixée dans la méthodologie à vingt minutes, l'atlas a finalement été carotté seulement sur une durée totale de 9min 42s car étant composé d'énoncés traduits ou reformulés, brefs (l'annexe I permet de s'en faire une bonne idée), sans blancs, il était trop long d'ajouter bout à bout divers énoncés jusqu'à obtenir la durée totale souhaitée, tout en conservant un caractère aléatoire de sélection. Les énoncés utilisés lors du carottage sont présentés en annexe I. Certains contextes morphosyntaxiques n'étaient pas encore présentés sur la plateforme lors du carottage, et certains contextes ont parfois été réorganisés. J'ai sélectionné aléatoirement un énoncé par contexte morphosyntaxique, en prenant le premier témoin pour le premier contexte morphosyntaxique, le second témoin pour le second contexte morphosyntaxique et ainsi de suite. Le nombre de témoins étant limité, j'ai dû repartir du premier témoin quatre fois avant d'épuiser tous les contextes morphosyntaxiques retenus sur le site de présentation de l'ALAVAL à cette époque. Lorsqu'aucune proposition n'était faite par un témoin, je prenais le témoin suivant, en essayant de revenir à ce témoin par la suite pour ne pas trop déséquilibrer les données, créant parfois de longs décalages (voir annexe I). J'aurais pu recommencer toute la procédure une deuxième fois, en prenant cette fois-ci le second témoin pour le premier contexte morphosyntaxique et ainsi de suite, mais j'ai décidé que le carottage de l'ALAVAL avait été assez chronophage et je me suis arrêtée là, afin de ne pas trop empiéter sur les objectifs de départ de cette thèse et pour que mes efforts restent proportionnels à l'intérêt scientifique de la tâche en question. L'ALAVAL étant presque équilibré hommes/femmes et entre les villages de recueil, les données carottées d'une telle façon sont globalement représentatives de l'atlas (à l'exception du dernier passage pour les villages).

Comme nous l'avons vu en 4.1.1.4, il n'est pas facile de travailler à partir des données de l'ALAVAL, car la plateforme de présentation en ligne n'a pas été conçue pour que les données soient exportées sous un autre format que le format PDF, ce qui n'est pas pratique. Par ailleurs, l'impossibilité d'accéder aux questions qui ont été posées en français ne permet pas, par exemple, de voir si l'enquêteur s'est laissé entraîner et a réalisé un pataquès en français lorsqu'il attendait un pataquès en francoprovençal, même si Andres Kristol (2018) m'affirmait dans une communication personnelle :

> Vous pouvez être sûre que lorsque nos témoins disent [siŋ ʒ ɔʃ] 'cinq ours', ils n'ont pas entendu « cinq-z-ours » en français. L'apparition d'un morphème pluriel préconsonantique « insolite » n'a donc jamais été induite par la langue

de travail, mais elle a pu être inhibée par leur bilinguisme, c'est-à-dire les normes du français.

La question de l'influence de la langue de travail pose question, mais n'oublions pas enfin, plus généralement, que le français est une langue dominante, qui a pu influencer le francoprovençal au cours de son existence. Berchtold et al. (2019a : 103, 115) indiquent par exemple, dans le contexte précis des liaisons observées après le pronom sujet de 2.PL, qu'« une comparaison avec les cartes de l'*ALF* [...] laisse cependant penser que les formes avec consonne de liaison ont massivement progressé depuis le début du XX^{e} siècle », et que les quelques réalisations de [iz] ou [z] pour le pronom sujet de 3.PL masculin « correspondent au français parlé familier ordinaire et sont sans doute influencées par celui-ci ».

J'ai relevé quatre pataquès lors de ce carottage :

- dans le contexte NUM + NOM (spécialement élicité dans le questionnaire de recueil de l'ALAVAL) : après *quatre* (exemple 130) et après *cinq* (exemple 131, comparé à 132). Berchtold et al. (2019b : 27) analysent cette consonne comme un morphème de pluriel.
- dans le contexte auxiliaire + participe passé (exemple 133). D'après Berchtold et al. (2019a : 145), il pourrait s'agir de « formes ayant agglutiné un [ʒ] initial d'origine analogique, propagé probablement à partir de la consonne de liaison dans des syntagmes du type [n ɛ̃ʒ y, no ʃɛ̃ʒ y] 'nous avons eu, nous sommes eu' ».
- dans le contexte de + NOM.PL (exemple 134). Berchtold et al. (2019a : 21-22) nous expliquent pourquoi il ne s'agit pas de *des* ici :

  En français, la forme *des* réunit trois fonctions. (1) Elle sert de pluriel aux formes de l'article indéfini singulier (*un*, *une*) pour des objets qui peuvent être comptés (sg. *manger une pomme*, pl. *manger des pommes*). (2) Elle sert de pluriel aux formes de l'article partitif singulier (*du*, *de la*) pour des objets dits « massifs » qui ne peuvent pas être comptés (sg. *manger de la polenta*, pl. *manger des épinards*). (3) Elle sert de pluriel aux formes de l'article défini singulier (*le*, *la*) précédé de la préposition *de* (sg. *je me souviens du grand-père*, pl. *je me souviens des grands-parents*). Les parlers du Valais central et de la Vallée d'Aoste, en revanche, distinguent formellement les cas (1) et (2) d'une part, et (3) de l'autre. [...] Pour (1) et (2), ils utilisent une simple préposition *de* [...] à la place du français *des* (on dit donc « *manger de pommes* » et « *manger d'épinards* »). Certains parlers de cette zone (Arbaz, Les Marécottes, Lourtier) utilisent aussi un simple [d] pour le pluriel de l'indéfini et du partitif pré-

vocalique (type « *manger d'oeufs* », comme au singulier « *donner d'eau* »). La majorité des dialectes a cependant développé une forme spécifique du pluriel indéfini/partitif devant voyelle, avec une consonne de liaison [...] qui sert de morphème pluriel prévocalique (type « *manger de z-oeufs* »). Pour (3) en revanche, tous les parlers du Valais central possèdent des formes spécifiques pour le pluriel défini précédé de *de* [...], correspondant au français *des*.

(130) [ˈkatrœː ʒ aˈɲo]
quatre agneaux
ContheyF : Il n'y a plus que quatre agneaux dans la bergerie.

(131) [ʃɪm ʒ ɔrʃ]
cinq ours
MiègeM : Je n'ai jamais vu cinq ours d'un coup.

(132) [siŋ ˈbiʃɛ]
cinq biches
ÉvolèneF : J'ai vu cinq biches fuir avec leurs petits (leurs faons).

(133) [də tseˈnɑwʊ n ɛ̃ ʒʏ plæ̃ˈto]
litt. 'du chanvre j'ai eu planté'
OrsièresF : Des … j'en ai eu plantés.

(134) [lɛ l ɑ̃ miˈtʏ də ʒ ɑbɹɔˈwɑ pɔ pɔ ɛ ˈɑtsɛ]
là ils ont mis de [z] abreuvoirs pour pour les vaches
SavièseM : Les vaches boivent à l'abreuvoir.

D'autres sandhis non étymologiques ne constituent pas des pataquès en francoprovençal : il s'agit de [k] (exemples 135, 136 et 137), [g] (exemples 138 et 139) et [p] (exemple 140).

(135) [ʏ fɑːt kʲe v ɑlʏˈʃa dryˈmɪk dœˈvã mjɛːˈnːɛ]
il faut que vous allier dormir avant minuit
St-JeanM : Il faut que vous vous couchiez avant minuit.

(136) [e ˈvjʊkə siŋ ˈbiʃɛ]
j'ai vu cinq biches
ÉvolèneF : J'ai vu cinq biches fuir avec leurs petits (leurs faons).

(137) [ɶ̃n aj la lɛteˈrɪk ɶ̃ pɔʁˈtav]
on allait à la laiterie on portait [...]

ChalaisF : D'habitude, on n'achetait pas le fromage. On le faisait soi-même (extrait).

(138) [ʃʊg la mã]
sur la main
MontanaM : Une coccinelle s'est posée sur sa paume.

(139) [ɔ̃ pɔᵡˈtav a la lɛteˈrɪg n ajɛ]
on portait à la laiterie on avait [...]
ChalaisF : D'habitude, on n'achetait pas le fromage. On le faisait soi-même (extrait).

(140) [aˈwe dɛ vɛˈlup ɛ dɛ ʁiˈbɑ̃]
avec des velours et des rubans
MontanaF : Le chapeau du costume féminin traditionnel (extrait).

(141) [tʊ saˈʁeɪ nʏtɔˈno]
tu serais étonné
ChamosonF : Si tu voyais l'oncle Antoine, tu serais étonné.

(142) [tʊ fʊrɛ tɔˈna]
tu serais étonné
IsérablesF : Si tu voyais l'oncle Antoine, tu serais étonné.

Ces consonnes sont dites *parasites* dans la dialectologie francoprovençale. Elles résulteraient de « diphtongue[s] étymologique[s] dont le deuxième élément s'est consonnifié, propagée[s] ensuite de manière analogique » (Berchtold et al., 2019b : 33). On voit qu'elles apparaissent après des voyelles fermées (/ɪ, ʊ, u/), et souvent devant initiale consonantique. Les consonnes impliquées, leur contexte d'apparition et l'expérience des dialectologues du francoprovençal sont autant d'arguments pour ne pas considérer ces sandhis externes non étymologiques comme des pataquès. Seul le dernier exemple (141, comparé à 142), pourrait en être un, le contexte, la consonne en question s'y prêtant, même si par précaution, je ne retiens pas cet exemple.

Par ailleurs, en francoprovençal, contrairement au français, il n'y a pas eu lexicalisation systématique d'un [t]- initial dans le mot *tante*, qui apparaît par exemple en liaison avec la forme à initiale vocalique (exemple 143).

(143) [mʊn ˈɑ̃ːta]
ma tante
Les MarécottesF : Ma tante mangeait toujours le jambon avec les pâtes.

En ce qui concerne les chiffres, comme en français, ils peuvent apparaître sans consonne finale devant initiale consonantique (exemples 132, 144 et 145). Cependant, devant initiale vocalique, il n'y a pas forcément de sandhi consonantique (pour 'neuf heures', cf exemples 146, 147, 148 et 149).

(144) [sa ˈmoθə]
sept mouches
IsérablesM : Avant l'orage, j'ai pu écraser sept mouches d'un coup.

(145) [wœ kɔˈpɛ̃]
huit copains
LourtierF : Nous étions huit amis pour faire cette farce.

(146) [nːœʊ̯ ˈœʏrə]
neuf heures
LiddesF : Comme neuf heures, je prends….

(147) [nyjœʀ]
neuf heures
Val-d'IlliezF : Je veux que les enfants se couchent avant neuf heures.

(148) [nɛ ˈɔjɛ]
neuf heures
TorgnonF : Je veux que les enfants se couchent avant neuf heures.

(149) [nʊ ˈɑwə]
neuf heures
BionazF : Je veux que les enfants se couchent avant neuf heures.

Ainsi, lors du carottage de l'ALAVAL, quatre pataquès ont été relevés. L'un d'eux (exemple 134) est spécifique à certaines variétés de francoprovençal. J'ai aussi mentionné l'existence de sandhis externes consonantiques non étymologiques dans cette langue, et j'en ai profité pour montrer que les phénomènes en marge de la liaison, comme le comportement des chiffres, était beaucoup moins systématique qu'en français standard. Le risque principal lors de l'étude des données de francoprovençal est de calquer nos connaissances de la liaison en français sur cette autre langue romane. Il est essentiel d'être très prudent, et il faudrait avoir une meilleure connaissance de la langue avant de se lancer dans une analyse plus poussée de la liaison. C'est pour cela, mais aussi parce que la fouille des données autrement que manuellement est impossible

avec la plateforme mise en place, que cette étude n'ira pas au-delà de cette phase de carottage. Tournons-nous à présent vers les résultats du carottage du corpus ALIPE.

4.1.3.2 Résultats du carottage du corpus ALIPE

Le corpus ALIPE a été carotté sur une durée totale de 19min 57s. J'avais au départ décidé de conserver sept minutes par enfant, en déduisant le temps de parole des parents. Finalement, j'ai analysé les productions enfantines et les productions des parents adressées à l'enfant, avec des blancs parfois assez longs, puisqu'il était chronophage et difficile de comptabiliser uniquement les productions enfantines. Un enregistrement pour chaque enfant a été aléatoirement choisi (manuellement ; il s'agit de **ali-baptiste-101229-2** pour Baptiste, de **ali-prune-091124-1** pour Prune et de **ali-salome-081117-2** pour Salomé), et les carottes ont été sélectionnées en divers endroits des transcriptions. J'ai modifié manuellement les annotations de la liaison dans les fichiers .cha selon un codage au format long (le système de codage de la liaison utilisé dans ALIPE est présenté en 2.1.3.4.2) en un codage au format court (le système de codage de la liaison adopté dans le cadre de cette thèse est présenté en 2.1.3.4.2), afin d'avoir une meilleure lisibilité.

Aucun pataquès n'a été noté lors du carottage. Dans les trois enregistrements sélectionnés dont étaient issus les extraits, une seule occurrence pourrait être analysée comme un pataquès, produite par une mère dans le contexte PRO sujet + PRO objet + [*l] + PRO objet *l'* + VER, et qui n'avait pas été annotée dans ALIPE :

(150) *MOT : tiens j(e) te [*l] l'ai ram(e)né. (ALIPE, ali-salome-081117-2)

De tels exemples ont été amplement discutés en 1.3.5.3.

Retenons donc qu'aucun pataquès n'a été relevé lors du carottage de l'ALAVAL, à l'exception d'une occurrence rencontrée dans les enregistrements d'où proviennent les extraits carottés. Cet unique pataquès, produit par une mère, peut paraître surprenant pour un corpus acquisitionnel. C'est pour cela que l'étape de première fouille permise par le corpus, déjà annoté pour la liaison, est essentielle. Pour l'heure, plongeons-nous dans les résultats des carottages des corpus ethnomusicologiques, en commençant par le FJD.

4.1.3.3 Résultats du carottage du FJD

Le fonds Jean Dumas a été carotté sur une durée totale de 20min 09s. En 2018, lors du carottage du fonds, celui-ci présentait quatre corpus. Je souhaitais carotter cinq minutes pour chaque corpus. J'ai sélectionné aléatoirement deux chansons par corpus, mais la durée totale des enregistrements ne s'élevait qu'à une quinzaine de minutes. Plutôt que de prendre une troisième chanson au hasard par corpus et de dépasser la durée totale souhaitée, j'ai décompté pour chacun le laps de temps nécessaire jusqu'à atteindre une durée de cinq minutes, et cherché à chaque fois une chanson qui remplissait ce critère. Un chant, déclamé et non chanté, a été exclu. Les identifiants des onze chansons retenues sont les suivants : JD_a060_31, JD_a062_17, JD_a063_33, JD_a137_10, JD_a137_28, JD_a160_10, JD_a165_47, JD_a168_07, JD_a205_14, JD_a350_02 et JD_a354_04. Les chansons ont été transcrites et annotées à l'aide de CLAN.

Ce carottage a été l'occasion de réfléchir au statut à accorder aux passages déclamés et non chantés des chansons, d'essayer de comprendre le plan de nommage (non uniformisé) qui avait été mis en place pour ce fonds ou encore de considérer les différentes transcriptions et annotations possibles de certains phénomènes comme le schwa, les schwas non étymologiques à l'intérieur de mots, les élisions, les (non)-réalisations de [l] à la finale des pronoms sujets *il(s)* et *elle(s)*, les liaisons à l'intérieur des numéraux, les existentiels, les vers bissés… pour lesquels j'ai tranché, pour la plupart, en 2.3.4.

Lors du carottage de ce fonds, j'ai relevé cinq pataquès :

(151) *SGR : je me suis rapproché d'elle qu'avez vous bell(e) [*z] à pleurer ?
(JD_a137_28, 63250, Lucienne Jailler)

(152) *SGR : tu mettras la main [*z] à la plume pour la première fois.
(JD_a137_28, 63250, Lucienne Jailler)

(153) *SGR : je dirai [*z] à mon.
*SGR : pèr(e) viens m(e) décharnasser !
*SGR : je dirai [*z] à ma mère fais moi d(e) la soup(e) au lait !
(JD_a165_47, 43800, Germaine Varenne)

(154) *SGR : j'y pleur(e) mon coeur en gage.
*SGR : il est bien [*z] en danger.
*SGR : il est bien [*z] en danger.
(JD_a168_07, 63550, Marie Rigodias)

(155) *SGR : ell(e) se mit [*z] à pleurer ell(e) se mit [*z] à pleurer.
[…]

*SGR : ell(e) se mit [*z] à chanter ell(e) se mit [*z] à chanter.
(JD_a168_07, 63550, Marie Rigodias)

J'y ai également observé une réalisation de [*n] dans une phrase affirmative (exemple 156), cas que j'aborderai plus en détail en présentant les résultats du carottage du fonds Charles Joisten en 4.1.3.4.

(156) *SGR : passant devant sa porte n'en demand(e) à loger.
*SGR : passant devant sa porte n'en demand(e) à loger.
(JD_a060_31, 43810, Virginie Granouillet)

Ce fonds présentait enfin une occurrence de liaison erratique qui aurait été considérée, selon son contexte d'apparition (nom singulier + invariable), comme « interdite » par Delattre (1947 : 153), comme dans l'exemple *le sapin ou le pin*, se rapprochant le plus de l'occurrence que j'ai relevée (exemple 157), qui lui servait d'illustration. En 157, on note deux réalisations de [n] après le nom commun *matin* et devant initiale vocalique, alors que le locuteur produit la voyelle orale [ɛ̃] devant initiale consonantique dans *vient*, sans consonne nasale finale.

(157) *SGR : le matin [n] et le soir.
*SGR : son amant la vient voir le matin [n] et le soir.
*SGR : son amant la vient voir.
(JD_a160_10, 43800, Jean-Pierre Arnaud)

Cinq pataquès ont de la sorte pu être notés lors du carottage du FJD. Le carottage de ce fonds m'a permis d'envisager différentes manières de transcrire et/ou d'annoter certains phénomènes. J'ai pour finir observé quelques curiosités comme une liaison habituellement considérée comme « interdite ». Venons-en aux résultats du carottage d'un autre corpus ethnomusicologique, le fonds Charles Joisten.

4.1.3.4 Résultats du carottage du fonds Charles Joisten

Le fonds Charles Joisten a été carotté sur une durée totale de 19min 31s. J'ai choisi aléatoirement des enregistrements qui devaient répondre aux caractéristiques suivantes : je cherchais des chansons, en français, de tradition orale, présentant un audio et interprétées par des chanteurs différents. Trois chansons de chanteurs déjà analysés ont par la suite été rajoutées pour se rapprocher de la durée totale prévue. Aucune chanson n'a semblé assez courte pour arriver à un total de vingt minutes. Les identifiants des onze

chansons retenues sont les suivants : M0996_SON57_1_2_A_01, M0996_SON57_1_2_A_03, M0996_SON57_2_3_A_04, M0996_SON57_2_5_A_03, M0996_SON57_2_6_A_05, M0996_SON57_3_1_A_15, M0996 _SON57_3_4_A_03, M0996_SON72_1_01, M0996_SON73_20_5_A_01, M0996_SON76_19 _A_01 et M0996_SON76_19_A_03. Ont été exclues les chansons catégorisées comme « chansons modernes (1900-1940) » dans le fonds. Sur ce point, Charles Joisten semble avoir été moins sélectif lors de sa collecte que Jean Dumas. Les chansons sélectionnées ont été transcrites et annotées à l'aide de CLAN.

Les chansons de tradition orale de ce fonds se sont révélées très semblables à celles du FJD (malgré une différence substratique évoquée en 4.1.1.5). Lors du carottage, j'ai notamment remarqué de nombreux doublons, avec peu d'inédits. Ceci serait intéressant pour travailler sur la transmission aréale de pataquès apparaissant dans les vers de chansons récurrentes, pour donner suite aux observations que j'avais formulées dans mon mémoire de master 2 (Coutanson, 2017). Comme cette orientation n'est plus mon objectif, ce fonds présente ici peu d'intérêt pour moi. C'est enfin des impératifs de temps qui m'ont fait renoncer à poursuivre l'étude d'un second fonds ethnomusicologique.

Lors du carottage de ce fonds, j'ai relevé deux pataquès :

(158) *SGR : sous le rocher y a [*t] une fontaine le rossignol y chante nuit et jou(r).
(Fonds Charles Joisten, M0996_SON76_19_A_03, Marcelle Sage)

(159) *SGR : n'est pas si bell(e) que toi mais (e)ll(e) est [p] un peu plus riche.
%pho : nɛ pa si bɛl kə twaː mɛ lɛ pɶ̃ pø ply ʁiʃə
(Fonds Charles Joisten, M0996_SON76_19_A_01, 38950, Marcelle Sage)

Ce dernier cas est ici comptabilisé comme un pataquès, bien que comme je l'avais mentionné en 1.3.5.1.1, il puisse s'agir d'un sandhi externe apparu par harmonie consonantique, sans lien avec le phénomène de liaison.

Il m'a été difficile de trancher si un cas particulier de sandhi externe inattendu était à considérer comme un pataquès : il s'agit des [*n] qui apparaissent dans des phrases affirmatives, comme dans les exemples 160, 161 et 162 issus du fonds Charles Joisten. Selon la définition adoptée en 1.3.2.4, tout pataquès est une liaison. Ici, le doute subsiste, puisque cette consonne peut provenir d'une métanalyse du pronom *en* en contexte de liaison (ex : *il en est revenu* pour l'exemple 160, avec agglutination du [n] au verbe *est*, chunk remobilisé par la suite séparément d'un pronom *en*), d'une métanalyse de la négation (ex : *il n'est pas revenu* pour l'exemple 160, avec agglutination du [n] sur le verbe *est*, chunk remobilisé par la suite dans une phrase affirmative) ou encore d'une allomorphie du pronom *en*, prononcé [nɑ̃], comme dans *Qui n'en veut ?* Enfin, ce phé-

nomène pourrait être favorisé par une voyelle nasale précédente, par assimilation progressive, comme l'illustrent les exemples 160 et 161, produits après [ɑ̃] ou être favorisé par le pronom *en* qui suit, par assimilation régressive, comme l'illustrent les exemples 156 p.220 et 162. Bien que la notation [*x] puisse noter tout sandhi externe inattendu, après avoir commencé à annoter le phénomène ici étudié [*n] dans le cadre de cette thèse, j'ai décidé d'abandonner cette notation pour ‹n'›. Sinon, de réels pataquès en [*n] comme dans l'exemple 48 p.102 se retrouvent noyés dans la masse des occurrences de ce phénomène. Considérer ou non ces exemples comme des pataquès n'est pas sans conséquence : l'ordonnancement des consonnes de pataquès (cf 1.3.5.1.1) notamment s'en trouve modifié, en jouant sur la part relative de la consonne /n/.

(160) *SGR : au bout de quinz(e) ans tout au plus le beau galant n'est revenu.
*SGR : s(i) il s'en va droit au logis du père en présentant tous ses respects.
(Fonds Charles Joisten, M0996_SON72_1_01, Marcel Simon)

(161) *SGR : en lui donnant la bagu(e) en or le beau galant n'est tombé mort.
(Fonds Charles Joisten, M0996_SON72_1_01, Marcel Simon)

(162) *SGR : la second(e) dans(e) qu'on fait Pierre n'en fit de même.
(Fonds Charles Joisten, M0996_SON76_19_A_01, Marcelle Sage)

Par ailleurs, j'ai relevé dans ce fonds une occurrence de liaison erratique qui aurait été considérée, selon son contexte d'apparition (le contexte nom propre + X), comme « interdite » par Delattre (1947 : 153), comme dans l'exemple *Jean espère venir*, se rapprochant le plus de l'occurrence que j'ai observée (exemple 163), qui lui servait d'illustration. En 163, un [n] est produit deux fois après *Françon* devant initiale vocalique, alors que le même prénom à la pause ne présente pas de consonne nasale finale.

(163) *SGR : donne moi des nouvelles de Françon !
*SGR : Françon [n] est bien malade.
*SGR : en danger de mourir.
*SGR : Françon [n] est bien malade se dit on.
(Fonds Charles Joisten, M0996_SON73_20_5_A_01, Émile Escalle)

Enfin, j'ai pu noter lors de ce carottage une réalisation de liaison en l'absence d'un M1, déjà présentée en 1.2.2.1.4 (exemple 20 p.37).

J'ai ainsi décompté deux pataquès lors du carottage du fonds Charles Joisten, fonds qui n'a pas été retenu pour une étude sur corpus, puisque les chansons collectées étaient

trop comparables à celles du FJD. La présentation des résultats de ce carottage a été l'occasion de mentionner la présence de *n'* dans des phrases affirmatives, cas que j'ai décidé d'écarter et de ne pas considérer comme un pataquès, ou encore une occurrence de liaison habituellement considérée comme « interdite ». Je terminerai en exposant les résultats du carottage du corpus PFC.

4.1.3.5 Résultats du carottage du corpus PFC

Le corpus PFC a été carotté sur une durée totale de 21min 04s. J'ai sélectionné aléatoirement quatre locuteurs, deux hommes et deux femmes issus de quatre zones francophones différentes : nord de la France, sud de la France, Canada et Afrique subsaharienne. Il s'agit de :

- 61ajb1, un homme de 70 ans ayant participé à l'enquête de Domfrontais (Normandie, France hexagonale)
- 12acr1, une femme de 26 ans ayant participé à l'enquête de Rodez (Aveyron, France hexagonale)
- cqejb1, un homme de 52 ans ayant participé à l'enquête de Trois-Rivières (Québec, Canada)
- maaaw1, une femme de 53 ans ayant participé à l'enquête de Bamako (Mali, Afrique subsaharienne)

J'ai opté pour les conversations, guidées ou libres (dans ce dernier cas, j'ai également pris en compte les interventions de leurs interlocuteurs). J'ai utilisé la fonction praat2chaat du logiciel CLAN afin de passer de documents au format .TextGrid (logiciel Praat ; cf Boersma & Weenink, 1992) à des documents au format .cha (logiciel CLAN ; cf. MacWhinney, 2000a). J'aurais également pu utiliser l'outil TeiCorpo (cf Modèles, Dynamiques, Corpus - UMR 7114 (MoDyCo), 2016 et Liégeois et al., 2015), dont je n'avais pas encore connaissance. J'ai utilisé la commande kwal afin de retirer les lignes dépendantes d'annotation de la liaison et du schwa (@Eg, @Bg), supprimé les ‹,› et remplacé les majuscules en début d'énoncés par des minuscules, renseigné les identités de chaque locuteur avant de relier leurs interventions respectives à leur code participant ou encore revu les chevauchements, intégrés dans les fichiers d'origine dans les tours de parole des locuteurs.

Lors du carottage de ce corpus, j'ai observé que le locuteur normand, 61ajb1, âgé de 70 ans, ne réalisait pas de liaisons après les pronoms sujets de 3.PL masculins *ils*, comme on peut le voir dans les exemples 39 p.95, 164 et 165. Ce phénomène est abordé en 1.3.2.3.

(164) *JB : [...] tes tes parents c'était comme ça Suzanne ?
*JB : ils essayaient [ilɛsɛjɛ] de vider les tonneaux [...].
(PFC, Domfrontais (Normandie, France, Europe), 2004, 61ajb1)

(165) *JB : [...] les restaurants les ils avaient [ilavɛ] des tonneaux chez eux dans leur cave.
(PFC, Domfrontais (Normandie, France, Europe), 2004, 61ajb1)

J'ai aussi relevé plusieurs exemples de [*l]. L'exemple 166, transcrit *C'était combien de litres qu'on l'avait le droit mais encore là dans l(e) temps ?* dans PFC, est ambigu : s'agit-il d'un double marquage de l'objet (à l'aide du pronom *l'* et du syntagme nominal *le droit*) ou est-ce que *l'* correspond à un pataquès en [*l], à la manière de ce qu'on observe en français canadien après le pronom *on* ? Ces cas avaient déjà été traités avec prudence par Morin (1982), et une harmonie consonantique n'est pas à exclure ici. En ce qui concerne l'exemple 167, il est transcrit *Tout ça ici* dans PFC. Les variétés canadiennes présentent de tels pataquès (voir 1.3.6.3). Le dernier exemple, 168, est transcrit *Ceux qui voyagent ils voyagent* dans PFC. Ce cas de figure a été longuement abordé en 1.3.5.3. J'ai considéré ces deux derniers cas comme des pataquès, malgré l'initiale consonantique du M2 dans le second cas. Il est très étonnant que ce soit ce type de pataquès, non annoté dans PFC[6] et avec une consonne de pataquès peu prototypique, notamment pour cette région francophone, qui soit apparu lors du carottage de ce corpus.

(166) *JB : c'était combien d(e) litres qu'on [*l] avait l(e) droit xxx dans l(e) temps ?
(PFC, Domfrontais (Normandie, France, Europe), 2004, 61ajb1)

(167) *SP1 : ceux qui veulent doivent voyager pour venir faire leur passeport ici.
*SP1 : tout ça [*l] ici.
(PFC, Bamako (Mali, Afrique subsaharienne), 2008, maaaw1)

(168) *SP1 : ceux qui [*l] voy(a)gent ils voy(a)gent.
(PFC, Bamako (Mali, Afrique subsaharienne), 2008, maaaw1)

De plus, j'avais déjà présenté une réalisation phonétique [ʒ] d'une liaison en /z/, par harmonie consonantique, relevée lors de ce carottage (exemple 12 p.30).

J'ai enfin rencontré deux exemples de liaison habituellement invariables ici non réalisées (exemples 169 et 170).

[6] La qualité audio de l'enregistrement dont sont issus ces deux exemples n'étant pas très bonne, on peut toutefois douter qu'ils aient réellement été produits.

(169) *CR1 : [...] elle a démissionné d(e) son [o*] autre travail [...].
(PFC, Rodez (Aveyron, France, Europe), 2002, 61ajb1)

(170) *SP1 : mais en tout cas on on demande à Dieu quan(d) même à ce qu'i(l) nous aide.
SP1 : nous ou- nous nous [o] ouvrir une porte comme on sait pas par.
*SP1 : par quel moyen.
(PFC, Bamako (Mali, Afrique subsaharienne), 2008, maaaw1)

Pour conclure, seuls deux pataquès semblent pouvoir être relevés lors du carottage du corpus PFC, peu prototypiques. Les extraits retenus révèlent aussi l'absence de réalisation de liaisons après le pronom *ils* pour un locuteur, une réalisation phonétique inhabituelle d'une liaison, par harmonie consonantique, ou encore des non-réalisations de liaisons habituellement considérées comme invariables.

Ainsi, dans les dix ou vingt minutes retenues lors de cette phase de carottage de l'ALAVAL, des corpus ALIPE et PFC et des fonds Jean Dumas et Charles Joisten, quelques pataquès ont été relevés, ce qui se révèle encourageant pour une étude sur corpus à venir. Seul le FJD sera étudié systématiquement, ainsi que le corpus de tops 1, projet de corpus ayant émergé après cette étape de carottage. On peut néanmoins douter de la pertinence d'une telle étape. Celle-ci m'a permis d'appréhender la transcription et l'annotation sous CLAN à venir et de borner toujours plus le phénomène de pataquès. J'ai en effet mentionné dans ce point certains phénomènes marginaux comme le comportement des numéraux devant initiale consonantique ou vocalique, les consonnes dites « parasitiques » en francoprovençal, les *n'* apparaissant dans des phrases affirmatives, des non-réalisations de liaisons habituellement considérées comme invariables, des réalisations de liaisons généralement considérées comme impossibles, des absences de M1 ou de réalisations de liaisons après le M1 *ils*, des [*l] équivoques pouvant correspondre à un double marquage de l'objet ou constituer des pataquès ou encore des liaisons dont la réalisation phonétique pourrait résulter d'une harmonie consonantique. Un carottage reste chronophage, pour n'obtenir que des résultats non représentatifs. Nous avions décidé d'expérimenter une telle phase dans le cadre de ce projet de thèse, mais on peut se demander si cela serait à réitérer pour des études postérieures.

4.1.4 Résultats de la première fouille des corpus

4.1.4.1 Résultats de la première fouille du corpus ALIPE

Le codage des liaisons d'ALIPE a permis une première fouille des données. Cette étape a nécessité environ huit heures de travail, notamment pour recopier et classifier tous les résultats. Les requêtes utilisées sous CLAN sont présentées en annexe II. Quelques exemples sont proposés ici à titre d'illustration, d'autres, redondants, sont disponibles en annexe III. J'ai observé :

- des agglutinations de la consonne de liaison sur le M2 apparaissant dans un tour de parole séparé du tour de parole contenant le M1 :

(171) *CHI : < il dit [^ Syntctx = I expecCons = t realCons = o obliOpt = 0][7] aux [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] oi–> [= ! chante] +...
*CHI : < [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] oiseaux z–> [= ! chante] +...
(ALIPE, ali-baptiste-110713)

(172) *FAT : ou des +...
%add : à CHI
*CHI : + + [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] arbres !
(ALIPE, ali-prune-071120-1)

(173) *FAT : un +...
%add : à CHI
*CHI : + + [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] hélicoptère !
(ALIPE, ali-prune-071122)

(174) *FAT : trois quoi ?
%add : à CHI
*CHI : [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] arbres !
(ALIPE, ali-prune-071120-1) x2

(175) *FAT : gros quoi ?
%add : à CHI
*CHI : + + [^ Syntctx = Z expecCons = o realCons = z obliOpt = 0] ourson !

[7] En Syntctx est annoté « le contexte syntaxique », en expecCons est indiquée « la consonne de liaison attendue » et en realCons « la consonne effectivement réalisée par le locuteur, ou l'absence de réalisation » ; enfin, en obliOpt est annoté « le caractère variable ou catégorique de la liaison » (Liégeois, 2014 : 183). Pour de plus amples explications du système de codage, voir Liégeois, 2014 : 183.

*FAT : hein ?
%add : à CHI
*CHI : + + [^ Syntctx = Z expecCons = o realCons = z obliOpt = 0] ourson !
(ALIPE, ali-prune-080714-2)

Je ne considère pas ces exemples comme des pataquès. En 171, on note une amorce chez Baptiste en contexte de liaison. La longueur de la pause doit être à l'origine de l'annotation en deux tours de paroles. Le reste des occurrences concerne Prune et son père. La plupart du temps, la fillette produit la liaison directement après que son père a produit le M1 (ne pas oublier que certains résultats sont présentés en annexe III). On peut ainsi noter dans ce corpus acquisitionnel des liaisons lorsque les mots 1 et 2 sont produits par des locuteurs différents. Parfois, comme en 174 et 175, un *quoi* ou *hein* sépare les mots 1 et 2. Terminons en précisant que de nombreuses occurrences de liaisons ou de pataquès dans les échanges entre Prune et son père relèvent d'élicitations suite à divers stimuli : images, devinettes... Il faut faire preuve de prudence à leur égard, ces pataquès ne pouvant être analysés au même titre que ceux réalisés spontanément par l'enfant.

- des agglutinations de la consonne de liaison sur le M2 sans lien avec le tour de parole précédent :

(176) *CHI : [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] a plus ça aussi ?
(ALIPE, ali-baptiste-101228-2)

(177) *CHI : ah tu [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] en ra– [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] a rach(e)té + .. ?
(ALIPE, ali-salome-090716)

(178) *CHI : i(l) [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] appuie sur un p(e)tit bouton ça s(e) lève.
(ALIPE, ali-prune-091125-1)

(179) *CHI : [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] en reste pas beaucoup faut qu'on [^ Syntctx = B expecCons = n realCons = n obliOpt = 1] en rachète.
(ALIPE, ali-salome-090716)

(180) *CHI : papa est-c(e) que tu peux faire le [^ Syntctx = Z expecCons = o realCons = l obliOpt = 2] l'éléphant ?
[...]

*CHI : et j(e) veux qu(e) tu fasses [^ Syntctx = I expecCons = z realCons = o obliOpt = 0] encore le [^ Syntctx = Z expecCons = o realCons = l obliOpt = 2] l'éléphant. (ALIPE, ali-salome-090710)

(181) *CHI : rega(r)d(e) moi il a renversé mon [^ Syntctx = A expecCons = n realCons = l obliOpt = 1] assiette xxx <là> [>]. (ALIPE, ali-baptiste-110707)

À nouveau, je ne considère pas ces exemples comme des pataquès. En 176 et 177 ([n] *a racheté*), on est face à un M1, *en*, absent (non reformulé en 177). Les exemples 177 (*n'en ra-*) et 178 nous rappellent le phénomène des *n'* apparaissant dans des phrases affirmatives mentionné en 4.1.3.4 lors des carottages du FJD et du fonds Charles Joisten. Les hypothèses que j'avais formulées restent d'actualité. On a d'autant plus envie d'être prudent que ce phénomène semble annoté *n'* ailleurs dans le corpus (cf par ex. 192 p.230). Liégeois (2014 : 386) semble pencher du côté d'un chunk issu de la métanalyse d'une négation, en commentant une production similaire de Baptiste comme suit :

> À l'oral, le *ne* étant particulièrement peu fréquent [...], c'est essentiellement l'adverbe *pas* ou *plus* qui porte le sens de la négation. Ainsi, en voulant produire [un] énoncé affirmatif [...], Baptiste se baserait sur cette construction en supprimant la forme à laquelle est rattachée le sens de négation[.]

Par la suite, il propose aussi une autre explication, en s'appuyant sur une occurrence produite par Prune, *il* [*n] *appuie* (Liégeois, 2014 : 404) :

> Les contextes de liaison verbaux étant à la fois récurrents et peu diversifiés lexicalement, ceux-ci resteraient mémorisés sous la forme d'expressions lexicalisées plus longtemps que les collocations nominales et seraient segmentées à un stade ultérieur [...]. Il s'agit [ici] de la collocation *il appuie*, au sein de laquelle la fillette insère la CL /n/ : [ilnapɥi]. Nous pouvons postuler que Prune a mémorisé [...] plusieurs exemplaires de la forme verbale, comme [napɥi] et [zapɥi] par exemple. Cette erreur [...] résulterait ainsi de l'insertion de l'exemplaire [napɥi] au sein de la construction [*il* + V].

L'exemplaire à [n] initial pourrait par exemple provenir d'une métanalyse du pronom personnel sujet *on*. En 179, nous sommes en présence d'une phrase négative, et il faudrait a minima considérer que c'est une double négation (*n'* ... *pas*), à

moins qu'il ne s'agisse à nouveau du même phénomène si l'enfant ne réalise pas de double négation ailleurs dans le corpus. Dans tous les cas, mieux vaudrait ne pas l'annoter comme une liaison. Pour les exemples 180 et 181, on peut adopter une analyse en faisant appel à la notion de chunk comme nous le verrons lors de l'analyse des pataquès. En 180, au déterminant défini *le* est adjoint le chunk *l'éléphant* [lelefɑ̃] et en 181, au déterminant possessif *mon* est adjoint le chunk *l'assiette* [lasjɛt]. Liégeois (2014 : 387) analyse la deuxième occurrence de la sorte :

> Nous interprétons cette erreur comme le fruit de l'insertion du mauvais exemplaire du Mot2 *assiette* dans le *slot* disponible de la construction |*mon* + X|. Cette interprétation est doublement motivée. Premièrement, si [...] le nom *assiette* montre une forte tendance à l'emploi au singulier [...], il est le plus souvent employé après l'article défini *le* dont l'élision obligatoire du schwa rend possible la mémorisation d'un exemplaire /lasjɛt/. [...] Cette force de la fréquence favoriserait donc la mémorisation de l'exemplaire et son ancrage en mémoire. De plus, les deux seules fois où Baptiste emploie correctement le nom *assiette*, il le fait précéder de la consonne /l/.

Mais comme ces sandhis inattendus ne sont pas liés au phénomène de liaison, je ne catégorise pas ces exemples comme des pataquès.

- de potentielles erreurs d'annotation :

(182) *FAT : des des [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] haricots ?
%add : à CHI
(ALIPE, ali-prune-071120-1)

(183) *FAT : < ben si > [<] tu sais on [^ Syntctx = B expecCons = n realCons = n obliOpt = 1] en [^ Syntctx = B expecCons = n realCons = n obliOpt = 1] a [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] en [^ Syntctx = B expecCons = n realCons = n obliOpt = 1] a parlé t(out) [^ Syntctx = C expecCons = t realCons = t obliOpt = 1] à l'heure.
(ALIPE, ali-prune-071122)

(184) *MOT : alors [^ Syntctx = J expecCons = z realCons = l obliOpt = 0] elle vous mon– elle vous mon– elle elle elle montre des ses elle elle vous montrait ses < xxx > [>] ses jouets.
(ALIPE, ali-salome-081120)

(185) *CHI : des [^ Syntctx = A expecCons = z realCons = n obliOpt = 1] histoires de

moi de ça ?
(ALIPE, ali-salome-090713)

(186) *CHI : elle va [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] en chercher ?
(ALIPE, ali-prune-071121-2)

(187) *CHI : < tu m(e) laves > [<] mes seins pendant c(e) temps là moi je m(e) lave euh mes [^ Syntctx = A expecCons = z realCons = n obliOpt = 1] oreilles.
(ALIPE, ali-salome-090710)

(188) *MOT : qu'est-c(e) que tu as [^ Syntctx = G expecCons = z realCons = t obliOpt = O] em– [^ Syntctx = G expecCons = z realCons = t obliOpt = 0] emm(e)né d'autre ?
(ALIPE, ali-salome-090716)

(189) *CHI : on va lui tirer les xxx [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] oreilles [^ Syntctx = F expecCons = z realCons = o obliOpt = 0] à maman.
(ALIPE, ali-prune-071125)

(190) *MOT : c'est des gamins [^ Syntctx = F expecCons = z realCons = o obliOpt = 0] en fait qui parlent dans en [^ Syntctx = Z expecCons = n realCons = n obliOpt = 1] intra-familial < dans > [>] la famille.
(ALIPE, ali-prune-091124-1)

(191) *CHI : < j(e) vais > [= ? vais] chercher [^ Syntctx = Z expecCons = ? realCons = n obliOpt = 2] aut(re).
(ALIPE, ali-salome-081119)

(192) *CHI : < pour demain > [<] i(l) n'en reste pour pour demain i(l) y en [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] en reste pour demain quand j(e) vais prend(re) mon bain pour mes p(e)tits baigneurs.
(ALIPE, ali-salome-090710)

(193) *CHI : bouh c'est ma [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] éléphant moi hein.
(ALIPE, ali-salome-081121)

(194) *CHI : de l'huile d'olives[8].
*CHI : moi j'aime bien le [^ Syntctx = A expecCons = 0 realCons = n obliOpt = 2] huile d'olive.
*FAT : l'huile d'olives.

[8] [sic]

%add : à CHI
(ALIPE, ali-salome-090716)

J'exclus aussi ces occurrences de la catégorie pataquès. Dans les exemples 182 et 183, la consonne de liaison [z] ne semble pas produite et en 184, aucun [l] ne semble réalisé. En 185, je doute que la consonne produite corresponde à un [n]. En 186, Prune semble produire plutôt *nous* [*o] *en chercher* et en 187, plutôt que *mes* [*n] *oreilles*, Salomé semble prononcer *mais non*. En 188, je transcrirais la production de la mère de Salomé de la sorte : *qu'est-ce que tu as t- t'as t'as emmené d'autre* plutôt que *qu'est-c(e) que tu as* [*t] *em-* [*t] *emm(e)né d'autre ?* Deux productions concurrentes semblent avoir été prévues, *qu'est-ce que tu as emmené ?* et *qu'est-ce que t'as emmené ?* Gardons tout de même à l'esprit qu'une occurrence comme *qu'est-ce que t'as* [*t] *emmené* ne serait pas incongrue pour un locuteur canadien (si la langue admet une telle variation en diatopie, en diachronie… l'éventualité d'une telle réalisation s'en trouve renforcée). En 189, *grosses* paraît produit en xxx devant *oreilles* (c'est aussi l'intuition de Liégeois, 2014 : 316), ce qui constituerait alors une liaison tout à fait attendue. En 190, le contexte syntaxique annoté (Z) semble erroné étant donné que la consonne produite, [n], concorde avec la consonne attendue. En 191, il semble que le mot *un(e)* soit réalisé devant *autre*. Le [n] ne serait alors pas une consonne de liaison. En 192, selon moi *il n'en reste* est produit deux fois. Par prudence, je n'annoterais pas la seconde production comme une liaison, puisque la première n'a pas été considérée comme telle. L'exemple 193 peut être analysé de deux façons : soit la prononciation de la voyelle du déterminant possessif *mon* n'est pas prototypique, soit le déterminant possessif féminin préconsonantique (*ma*) a été choisi au lieu de son équivalent prévocalique (*mon*). Dans tous les cas ce n'est pas vraiment la consonne de liaison produite qui pose problème, j'hésite donc à considérer une telle occurrence comme un pataquès, bien que normalement, selon ma définition, toute consonne de liaison inattendue par rapport à la consonne graphique finale du M1 constitue un pataquès. Enfin, en 194, Salomé semble produire [ləlqivdɔliv] plutôt que *le* [*n] *huile d'olive*.

- des occurrences que j'analyserais autrement :

(195) *CHI : nan xxx les [ˆ Syntctx = Z expecCons = o realCons = z obliOpt = 2] (j)ambes.
[…]
*CHI : ma [ˆ Syntctx = Z expecCons = o realCons = z obliOpt = 2] jambe.
[…]
*CHI : xxx des [ˆ Syntctx = Z expecCons = o realCons = z obliOpt = 2] (j)ambes.

[...]
*CHI : ju(s)te une tâche ?
(ALIPE, ali-baptiste-110101-1)

(196) *FAT : euh des [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] haricots.
%add : à CHI
(ALIPE, ali-prune-071122)

(197) *FAT : et lui c'est qui ?
%add : à CHI
*CHI : un [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] hibou !
*FAT : ouais.
%add : à CHI
*FAT : et s'i(l) y en [^ Syntctx = B expecCons = n realCons = n obliOpt = 1] avait plusieurs ça f(e)rait quoi +.. ?
%add : à CHI
*FAT : des +...
%add : à CHI
*CHI : des [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] hiboux.
[...]
%add : à CHI
*CHI : des [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] hiboux.
*FAT : ouais.
%add : à CHI
*FAT : et s'il était gros ça s(e)rait [^ Syntctx = H expecCons = t realCons = o obliOpt = 0] un gros +...
%add : à CHI
*CHI : un gros [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] hibou.
*FAT : un quoi ?
%add : à CHI
*CHI : un gros [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] hibou.
*FAT : et s'il était petit ?
%add : à CHI
*CHI : xxx +...
*CHI : +, < un gros [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] hibou > [= ! chuchote].
*FAT : d'accord un gros [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2]

hibou et s'il était petit +..?
%add: à CHI
(ALIPE, ali-prune-071124)

(198) *CHI: avant [^ Syntctx=J expecCons=t realCons=o obliOpt=0] i(l) y a les [^ Syntctx=Z expecCons=o realCons=z obliOpt=2] avec zizi.
[...]
*CHI: et les [^ Syntctx=2 expecCons=o realCons=z obliOpt=2] avec zizi.
(ALIPE, ali-prune-071120-1)

(199) *CHI: un [^ Syntctx=P expecCons=o realCons=n obliOpt=2] R@l.
(ALIPE, ali-prune-091126)

Dans l'exemple 195, on voit par deux fois que l'enfant prononce le mot *jambe* [zɑ̃b] au pluriel, mais aussi au singulier. Que le mot *juste* soit prononcé dans le même enregistrement avec un [z] initial finit de nous convaincre que l'enfant n'a pas tout à fait acquis la prononciation du son /ʒ/, et je ne considère ainsi pas ces occurrences comme des pataquès (de la même façon que le fait Liégeois, 2014 : 316). Lorsqu'une liaison est produite avant des mots historiquement disjonctifs, comme en 196 pour le mot *haricots* ou en 197 pour le mot *hibou(x)*, cela ne constitue pas selon moi des pataquès (c'est également ce que pense Liégeois, 2014 : 315) mais une évolution attestée de la langue. Si le père de Prune lui-même produit des liaisons avec ces mots, on ne sera pas surpris que sa fille les réalise également. De la même façon, on s'attend à ce que le contexte de liaison de 198 soit disjonctif. Mais ici, Prune choisit de faire la liaison et produit une liaison erratique, que la définition que j'ai adoptée de la liaison n'empêche pas de considérer comme telle. Enfin, il en va de même en 199, où une liaison est produite devant le nom de la lettre ‹r›, sans disjonction donc.

- des occurrences pour lesquelles il est impossible de trancher s'il s'agit réellement de pataquès :

(200) *CHI: <tu es [^ Syntctx=H expecCons=z realCons=t obliOpt=0] un> [=? tu éteins] xxx.
(ALIPE, ali-prune-091124-1)

(201) *CHI: c'est [^ Syntctx=Z expecCons=o realCons=n obliOpt=2] e(s)cargot vois p(l)us rien vois pas xxx <voit> [>].
(ALIPE, ali-salome-081117-1)

En 200, l'annotateur a hésité entre *tu es* [*t] *un* et *tu éteins*. En 201, il y a un bruit

lors de la réalisation de la consonne, et il n'est pas possible de savoir si c'est un [*n] ou un *l'* qui est produit.

- enfin, des exemples que je considère comme des pataquès :

(202) *MOT : tiens donne moi [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] en [^ Syntctx = J expecCons = n realCons = o obliOpt = 0] une.
(ALIPE, ali-salome-081118)

(203) *MOT : quand tu peux prendre la personne entre quat(re) [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] yeux [^ Syntctx = F expecCons = z realCons = o obliOpt = 0] et lui c(e) que tu t(u) as dit là et ben écoute j(e) suis pas d'accord < avec toi > [>].
(ALIPE, ali-prune-091123-3)

(204) *FAT : un gros [^ Syntctx = E expecCons = z realCons = t obliOpt = 0] abricot +..?
(ALIPE, ali-prune-071124)

(205) *FAT : bon oh ben elle t'en [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] r(e)na elle t'a r(e)donné encore +...
(ALIPE, ali-salome-090713)

(206) *CHI : et elle elle a pas d(e) [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] oreilles !
(ALIPE, ali-prune-071124)

(207) *CHI : et les mammouths ne [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] ont pas de zizi.
[...]
*CHI : c'est les ceux ceux qui [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] ont des des des zizis !
(ALIPE, ali-prune-071120-1)

(208) *CHI : +, des [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] amis qui [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] étaient [^ Syntctx = H expecCons = t realCons = o obliOpt = 0] allés +...
(ALIPE, ali-prune-071124)

(209) *FAT : un gros [^ Syntctx = E expecCons = z realCons = o obliOpt = 0] ours (.) et un petit +...
%add : à CHI

*CHI : + + [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] ours !
(ALIPE, ali-prune-071124)

(210) *FAT : on n'est pas d'accord on n'a pas les mêmes [^ Syntctx = P expecCons = z realCons = z obliOpt = 0] idées.
[...]
*CHI : tous les deux vous [^ Syntctx = B expecCons = z realCons = z obliOpt = 1] avez pas la même [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] idée ?
(ALIPE, ali-prune-071120-2)

(211) *CHI : l'autre [^ Syntctx = E expecCons = o realCons = n obliOpt = 2] oeil il est coincé.
(ALIPE, ali-salome-090713)

(212) *CHI : et xxx i(l) peut fai(r)e < quoi > [= ? toi] le [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] oiseau ?
(ALIPE, ali-baptiste-110801)

(213) *FAT : ça c'est [^ Syntctx = H expecCons = t realCons = o obliOpt = 0] un + ...
%add : à CHI
*CHI : + + [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] orteil !
*CHI : deux [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] orteils un [^ Syntctx = A expecCons = n realCons = z obliOpt = 1] orteil.
(ALIPE, ali-prune-071120-1)

(214) *CHI : c'est mais ça c'est le [^ Syntctx = Z expecCons = 0 realCons = n obliOpt = 2] escargot xxx.
(ALIPE, ali-salome-090712)

(215) *CHI : oh ta [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] orange elle est tombée regarde.
FAT : < ma [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] orange > [>]
%add : à CHI
*CHI : < xxx > [<].
*CHI : < regarde > [>] !
*FAT : < mon [^ Syntctx = A expecCons = n realCons = n obliOpt = 1] orange > [<] elle est tombée ?
%add : à CHI
(ALIPE, ali-prune-071120-1)

(216) *CHI : tu sais comment [^ Syntctx = J expecCons = t realCons = o obliOpt = 0] on

xxx not(re) [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] yeux i(l) peut bouger ?
*FAT : notre quoi ?
%add : à CHI
*CHI : notre notre euh notre rond qui peut bouger dans les [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] yeux pas vrai qu'i(l) peut bouger ?
*FAT : notre oeil ?
%add : à CHI
(ALIPE, ali-prune-080715-2)

(217) *CHI : dans ce [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] avion !
(ALIPE, ali-prune-071122)

(218) *CHI : Mireille [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] abeille !
*FAT : ouais et ça s'appelle une +...
%add : à CHI
*CHI : + + [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] abeille !
(ALIPE, ali-prune-071122)

(219) *CHI : non c'est le papa de [^ Syntctx = Z expecCons = o realCons = t obliOpt = 2] ours brun !
*FAT : c'est la[9] papa de [^ Syntctx = Z expecCons = o realCons = t obliOpt = 2] ours brun ?
%add : à CHI
*CHI : oui.
*CHI : oh !
*FAT : et il est où t– euh petit [^ Syntctx = P expecCons = t realCons = t obliOpt = 0] ours brun alors ?
(ALIPE, ali-prune-071120-1)

(220) *CHI : +, le pied [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] ici +...
(ALIPE, ali-prune-071122)

(221) *CHI : sinon moi j(e) te transforme en [^ Syntctx = P expecCons = n realCons = z obliOpt = 1] étoile !
*FAT : ah mais comment tu vas faire ?
%add : à CHI
*CHI : je prends xxx transforme en [^ Syntctx = P expecCons = n realCons = z

[9] [sic]

obliOpt = 1] étoile !
(ALIPE, ali-prune-071120-2)

(222) *CHI : des petits euh (.) [^ Syntctx = E expecCons = o realCons = z obliOpt = 0] ananas.
(ALIPE, ali-prune-071120-1)

- Les parents produisent des pataquès dans ce corpus, comme les mères de Prune et de Salomé dans le contexte (VER.PRS.IMP.2.SG) + enclitique + enclitique, en 66 p.117 et 202. La mère de Prune utilise aussi l'expression *entre quatre* [*z] *yeux* en 203, réalisation qui est généralement considérée comme un pataquès normatif, n'ayant pas d'équivalent non normatif. À nouveau, si les parents font de tels pataquès, on ne s'étonnera pas que les enfants les produisent à leur tour en-dehors des données enregistrées dans le cadre de la constitution de ce corpus. En 204, le père de Prune produit un pataquès dans le contexte ADJ + NOM. Celui-ci utilisant régulièrement l'élicitation en contexte de liaison, on ne sait pas s'il est ici en train de tester sa fille afin de savoir si celle-ci accepte cette réalisation. On restera donc prudents vis-à-vis de cette occurrence. Les parents répètent parfois des pataquès produits par leurs enfants, comme on peut le voir en 215, 219, 336, 342 et 347. Ils apportent seulement occasionnellement une correction, comme en 215, 216 ou 337, ou reformulent, comme en 219 ou 332. En 205, le père de Salomé produit encore une occurrence complexe à analyser, que je transcrirais *bon oh ben elle t'en r(e)-* [*n] *a elle t'a r(e)donné encore*. Ici, deux productions semblent entrer en concurrence lors de son intervention : *elle t'en a r(e)donné* et *elle t'a r(e)donné encore x*, avec anaphore pronominale ou nominale à venir. Il est assez curieux que dans un premier temps, le préfixe r(e)- puisse se placer avant la liaison [n]. Ce locuteur se rend peut-être compte d'un potentiel problème d'interprétation en reformulant sa pensée.
  Certains pataquès des enfants sont par ailleurs semblables à ceux observés chez des adultes. Les pataquès après des *de* (comme en 206) tout d'abord rappellent un cas apparenté en francoprovençal (voir 4.1.3.1). En français, *de* et *des* peuvent être en concurrence au pluriel, le premier étant élidé devant initiale vocalique : *d'oreilles* vs *des oreilles*. C'est également le cas des pataquès dans le contexte complémenteur *qui* + VER.3.PL (comme en 207 (*qui* [*z] *ont*) et 208), cas exposé en 1.3.5.3. Chez les enfants, ce dernier cas s'étend ici après la négation en 207 (*ne* [*z] *ont pas*). Pour ces deux derniers cas, Liégeois (2014 : 394-395) propose l'explication suivante :

> [...] au cours de son développement, l'enfant va être confronté à davantage de collocations *ils* + Verbe l'amenant [...] à repérer progressivement l'analogie entre les différentes collocations rencontrées en input. [...] le lexique utilisé en [discours adressé à l'enfant] se diversifie, entraînant ainsi une multiplication des formes de Mots2 [...]. De ce fait, une construction plus abstraite se trouve renforcée et davantage ancrée en mémoire, ce qui entraîne [...] une segmentation de la construction lexicalisée sur la base du schéma standard de la syllabe en CV. Il découle de ce découpage la mémorisation d'exemplaires comme *z-ont* ou *z-étaient*. Ces exemplaires verbaux [...] peuvent être insérés dans un *slot* disponible pour une forme verbale. Il en résulte donc des erreurs d'adjonction [...][, p]ar exemple, [...] l'exemplaire *z-ont* [est inséré] dans le *slot* disponible pour un verbe dans la construction |N + *ne* + V + *pas*|. Cette hypothèse se trouve renforcée par la grande fréquence d'emploi de la CL /z/ avant les formes verbales *ont* et *étaient*.

Enfin, on trouve des pataquès dans le contexte ADJ + NOM, comme en 209, 210 ou 211. Ce corpus de Loïc Liégeois, plutôt qu'une collection d'exemples, se prête bien à une analyse en termes de constructions. On voit par exemple en 209 que [zuʁs] est proposé par l'enfant après *petit* à la suite de *gros ours* [gʁozuʁs]. Il en va de même pour 210, où le père produit la cooccurrence *mêmes* [z] *idées*, qui est reprise telle quelle au singulier par Prune, sous forme de chunk : *même* [*z] *idée*. D'autres pataquès semblent exclusivement enfantins. Ce seront tout d'abord les pataquès qui apparaissent dans le contexte DET + NOM, comme en 212, 213, 214 ou 218 (*une*... [*z] *abeille*). En 212, 213 et 218, la liaison du pluriel [z] est agglutinée au M2 dans un contexte singulier. En 214, la consonne de l'indéfini singulier, [n], est agglutinée au M2 dans un contexte défini singulier. De la même façon, les pataquès dans le contexte DET possessif + NOM, comme en 215 ou 216, semblent propres aux enfants. En 216, le chunk [zjø], avec le M2 pluriel *yeux*, est convoqué au lieu d'*oeil*. Ce type de pataquès apparaît enfin dans le contexte DET démonstratif + NOM, comme en 217. Un jeu fait du bruit, mais le pataquès ici produit semble plutôt être en [*z] que [*n]. En remobilisant à nouveau la notion, d'autres chunks sont révélés en 218 et 219. En 218, l'anthroponyme *Mireille l'abeille* a été décomposé, et un chunk que Prune utilise ailleurs dans la langue (toujours 218), [zabɛj], est venu remplacer [labɛj]. En 219, le chunk [tuʁsəbʁɛ̃] a été extrait de l'anthroponyme *Petit Ours Brun*. Pour continuer, on trouve une occurrence, 220,

dans le contexte NOM + ADV et une autre, 221, dans le contexte PRP + NOM. Pour terminer, la liaison peut apparaître après une hésitation (*euh*), comme on le voit en 222.

Un bilan quantitatif des occurrences de pataquès du corpus ALIPE est présenté dans le tableau 4.5. On note que globalement, les pataquès y sont rares, puisqu'on ne les rencontre que dans 0,87% (66/7598[10]) des contextes de liaison et de pataquès[11]. Rappelons-nous que ces résultats dépendent des pataquès retenus comme tels, que certaines données du corpus peuvent rester silencieuses et que de nombreuses réalisations de Prune semblaient consciemment élicitées par son père. Dans ce corpus acquisitionnel, les enfants produisent plus de pataquès (2,4% ; 55/2296) que les parents (0,2% ; 11/5302), et la différence se creuse encore davantage en excluant les données parentales de reprise de pataquès enfantins, pour atteindre 0,11% de réalisation (6/5302).

Enfants		**Parents**		**Parents hors reprises**		**TOTAL**	
%	n	%	n	%	n	%	n
2,40	55/2296	0,21	11/5302	0,11	6/5302	0,87	66/7598

TABLEAU 4.5 : Bilan quantitatif de la première fouille du corpus ALIPE (modifié depuis Coutanson, 2020 : 10 après réanalyse)

Ainsi, lors de cette première fouille du corpus ALIPE, nous pouvons éliminer le bruit, mais le silence ne peut être évité. L'annotation Z, ou liaison hors contexte, m'a permis d'obtenir la plupart des occurrences, mais j'ai également usé d'autres requêtes afin de parvenir à mes fins. Certaines réalisations ont été écartées, soit parce que je conteste l'annotation retenue, soit car elles ne correspondaient pas à ma définition du phénomène de pataquès : les liaisons apparaissent dans des tours de parole différents, le M1 est absent, il s'agit du phénomène de *n'* dans des phrases affirmatives, la consonne

[10] Le nombre total de contextes de liaison et de pataquès est obtenu en faisant la somme des contextes produits par les pères (2490), obtenus à l'aide de la requête CLAN `combo +s"[^ Syntctx=* expecCons=* realCons=* obliOpt=*]" @ +u +t*FAT`, par les mères (2812), obtenus à l'aide de la requête `combo +s"[^ Syntctx=* expecCons=* realCons=* obliOpt=*]" @ +u +t*MOT` et par les enfants (2296), obtenus à l'aide de la requête `combo +s"[^ Syntctx=* expecCons=* realCons=* obliOpt=*]" @ +u +t*CHI`, somme qui peut être vérifiée à l'aide de la requête `combo +s"[^ Syntctx=* expecCons=* realCons=* obliOpt=*]" @ +u`.

[11] Je parle de *contextes de liaison et de pataquès* afin d'inclure dans l'effectif total les contextes où des pataquès hors contexte de liaison sont apparus. Cela peut surprendre ; de tels contextes ont pourtant été annotés pour la liaison dans les corpus ALIPE ou encore PFC, bien qu'ils ne constituent pas des contextes de liaison.

initiale est remplacée par une autre car le phonème n'est pas encore acquis, la disjonction attendue n'est pas produite, etc… Cette première fouille du corpus s'est révélée fructueuse bien que chronophage, celui-ci présentant des pataquès enfantins, numériquement majoritaires, mais aussi quelques pataquès produits par des adultes, dont une bonne partie consiste en une reprise de pataquès enfantins. On gardera néanmoins à l'esprit que les activités centrées sur le langage comme les dénominations d'images ou les devinettes ont pu déclencher davantage de pataquès chez Prune. Les pataquès qui semblent purement enfantins seraient produits préférablement en contexte prénominal (cf aussi Liégeois, 2014 : 318). Cette première fouille s'est donc révélée cruciale, puisqu'avec l'étape de carottage seule, on aurait pu croire que ce corpus n'était pas très pertinent, étant donné que les échantillons retenus ne présentaient aucun pataquès. Passons dès lors aux résultats obtenus pour le corpus PFC.

4.1.4.2 Résultats de la première fouille du corpus PFC

Le codage des liaisons de PFC a permis lui aussi une première fouille des données. Cette étape a nécessité environ une heure et quart de travail. Marie-Hélène Côté a tenté, pour son article *La liaison en diatopie : esquisse d'une typologie* (Côté, 2017), de trouver des liaisons non standard dans le corpus PFC. Elle s'est vite heurtée à l'annotation « épenthétique » :

> […] [I]l convient de s'attarder au codage PFC, qui indique les liaisons dites 'épenthétiques'. Le protocole PFC définit ces liaisons comme n'ayant pas d'origine graphique, par exemple *il va à Paris* prononcé [ilvatapari] (Durand et al., 2009 : 34). Mais ce codage s'avère ambigu et ne permet que partiellement l'identification des liaisons non standard. D'une part, s'il est clair que le codage épenthétique s'applique aux cas d'insertion, inclut-il également les cas de substitution ? Le traitement de ces cas s'avère variable. […] le [t] qui accompagne *suis* est considéré comme épenthétique dans l'enquête de Trois-Rivières, mais pas dans celle de l'Université Laval.
>
> D'autre part, les liaisons non standard ne sont pas toutes reconnues comme telles, et donc codées. Par exemple, des formes laurentiennes comme *ça* [l] *arrive* sont à analyser comme contenant un [l] de liaison […]. Elles sont pourtant spontanément transcrites *ça l'arrive*. […]

L'annotation « liaison épenthétique » n'a ainsi pas été utilisée pour tous les pataquès, quand certains n'ont pas même été annotés lors du codage de la liaison. Les requêtes

auxquelles j'ai fait appel lors de cette première fouille du corpus sont présentées en annexe IV. Celles-ci sont vraiment dépendantes de notre connaissance préalable du sujet d'étude. La liste qui suivra n'est ainsi pas exhaustive et des pataquès annotés pourraient ne pas avoir été relevés. Connaître le codage de PFC m'a permis d'émettre la requête principale et ainsi d'accéder aux liaisons codées comme épenthétiques, puisque ce choix n'est pas proposé sur l'interface de requête en ligne[12], comme on le voit dans la figure 4.2.

FIGURE 4.2 : Possibilités de requêtes concernant le type de liaison sur l'interface de recherche de PFC

Les résultats à cette requête principale sont présentés en annexe V. On peut relever que 35/105 occurrences correspondent réellement à des pataquès (notons que de nombreuses annotations pour une locutrice de l'enquête d'Abidjan semblent être celles du schwa plutôt que de la liaison). Ces occurrences permettent d'orienter notre fouille vers d'autres requêtes, que ce soit des contextes lexicaux spécifiques pour le test de lecture ou des contextes lexicaux dans les conversations. L'annotation de PFC n'est pas conçue, contrairement à celle d'ALIPE, pour révéler une contradiction entre consonne de liaison attendue et consonne de liaison réalisée. Le logiciel Dolmen pourrait peut-être permettre une telle requête, en utilisant des expressions régulières et en se fondant sur les derniers caractères des M1, ce que je n'ai pas expérimenté. Voici les pataquès que j'ai relevés :

- des pataquès produits lors de la tâche de lecture du texte, dans les contextes suivants[13] :

(223) *a_en*, où trois locuteurs ont produit un [*t] (*Le Premier Ministre a* [*t] *en effet décidé* [...])

(224) *grand_émoi* et *grand_honneur*, où respectivement onze locuteurs et un locuteur ont produit un [*n] ([...] *en grand* [*n] *émoi*, [...] *ce grand* [*n] *honneur* ; cinq sont locuteurs de variétés américaines ou africaines, comme on le voit dans la

[12]Le codage de PFC est le suivant : liaison absente = 0, liaison enchaînée = 1, liaison incertaine = 2, liaison non enchaînée = 3, liaison épenthétique = 4, codage qui semble avoir été modifié depuis Durand et al., 2002.

[13]Les occurrences ont été vérifiées, mais l'ensemble des productions en lecture n'a pas été réécouté de façon systématique au vu du nombre de contextes et de réalisations.

figure 4.3 pour *en grand* [*n] *émoi*[14])

(225) *qu'il_appelle*, où deux locuteurs ont produit un [*z] ([…] *ce qu'il* [*z] *appelle* […] ; pluralisant oralement le syntagme)

(226) *mène_au*, où onze locuteurs ont produit un [*t] ([…] *qui mène* [*t] *au village* […] ; dix sont locuteurs de variétés européennes, comme on le voit dans la figure 4.4[14])

(227) *quatrième_aux*, où deux locuteurs ont produit un [*z] ([…] quatrième [*z] aux Jeux Olympiques […])

(228) *son_usine*, où deux locuteurs ont produits un [*z], probablement par harmonie consonantique, avant réparation (rcamp1 : […] *son* [*z] *son* [o] *usine* […], rcark1 : […] *son* [*z] *u- son* [n] *usine* […])

(229) *provoquer_une*, où une locutrice a produit un [*t], avant réparation (75xlv1 : […] *provoquer* [*t] *u-* [o] *une explosion*)

- des pataquès produits dans les conversations libres ou guidées :

(230) FV1 : Pis généralement quand je suis11t en congé ben je vais dessiner je vais lire pis euh,
(PFC, Saguenay (Canada, Amérique du Nord), 2006, cqbfv1)

(231) GT1 : Me suis11t en revenue en cinquante-deux.
(PFC, Saguenay (Canada, Amérique du Nord), 2006, cqbgt1)

(232) MC : Je suis14t allé euh,
(PFC, Trois-Rivières (Canada, Amérique du Nord), 2014, cqemc1)

(233) MC : Je te dirais vraiment l'espagnol quand je suis14t arrivé là-bas c'est, c'est, c'est ce qui l'a fait développer là.
(PFC, Trois-Rivières (Canada, Amérique du Nord), 2014, cqemc1)

(234) MC : Mais là j/ j/, je suis11t un peu perdu dans mes11z affaires de géo/-logie, pis de géographie là, mais me semble que, tu sais euh, d/ s/.
(PFC, Trois-Rivières (Canada, Amérique du Nord), 2014, cqemc1)

[14]Deux logiciels de cartographie ont été utilisés afin d'obtenir cette carte : les points d'enquêtes ont d'abord été ajoutés sur le logiciel open source uMap (2023), fondé sur OpenStreetMap, avant d'être exportés au format .kml puis importés sur le logiciel libre et gratuit Magrit (Viry et al., 2023), qui a permis d'obtenir la visualisation souhaitée (fonds de carte *Pays du monde (Polygones)* et visualisation *Natural Earth II*).

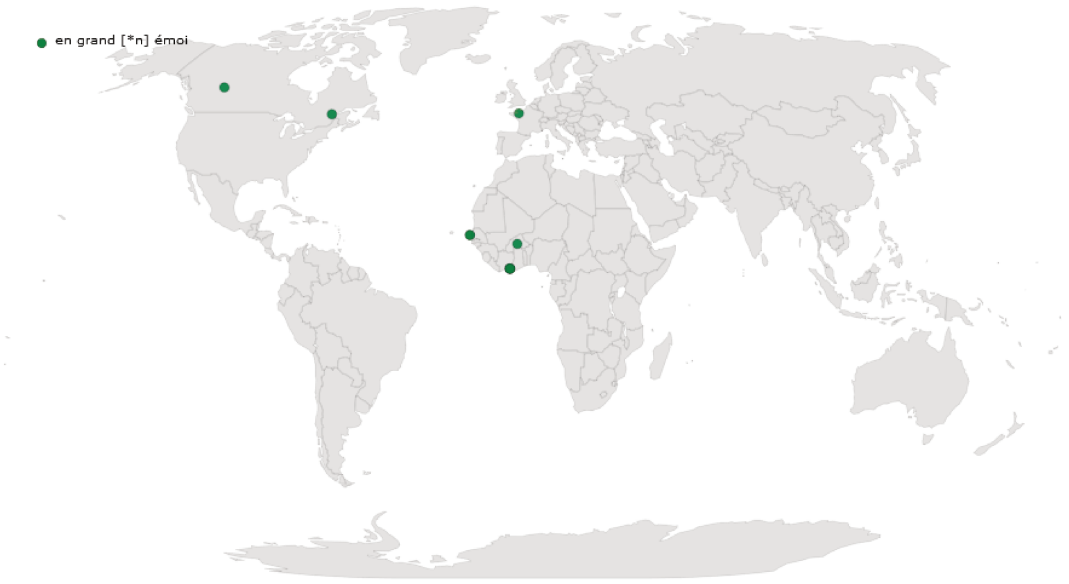


FIGURE 4.3 : Enquêtes de PFC où le pataquès *en grand* [*n] *émoi* a été produit dans la tâche de lecture du texte

(235) WD : Pis10 après ça je suis14t allé g/ euh, Ninstitution.
(PFC, Trois-Rivières (Canada, Amérique du Nord), 2014, cqewd1)

(236) WD : Alors je suis14t allé faire mon cours, pis je suis rentré aux services20 après de,
(PFC, Trois-Rivières (Canada, Amérique du Nord), 2014, cqewd1)

(237) WD : Il y avait20 un problème à Nlieu, je suis14t allé ouvrir,
(PFC, Trois-Rivières (Canada, Amérique du Nord), 2014, cqewd1)

(238) WD : quand je suis14t entré en cinquante-neuf,
(PFC, Trois-Rivières (Canada, Amérique du Nord), 2014, cqewd1)

(239) LC : Et pis moi ben j/, je suis14t allée à l'école, comme tout le monde.
(PFC, Trois-Rivières (Canada, Amérique du Nord), 2014, cqelc1)

(240) JB : absolument euh, moi je suis11t un, un gars de char, hein ?
(PFC, Trois-Rivières (Canada, Amérique du Nord), 2014, cqejb1)

(241) LL : 'non, faut tout le temps que je change, je suis14t un vrai artiste'.
(PFC, Trois-Rivières (Canada, Amérique du Nord), 2014, cqell1)

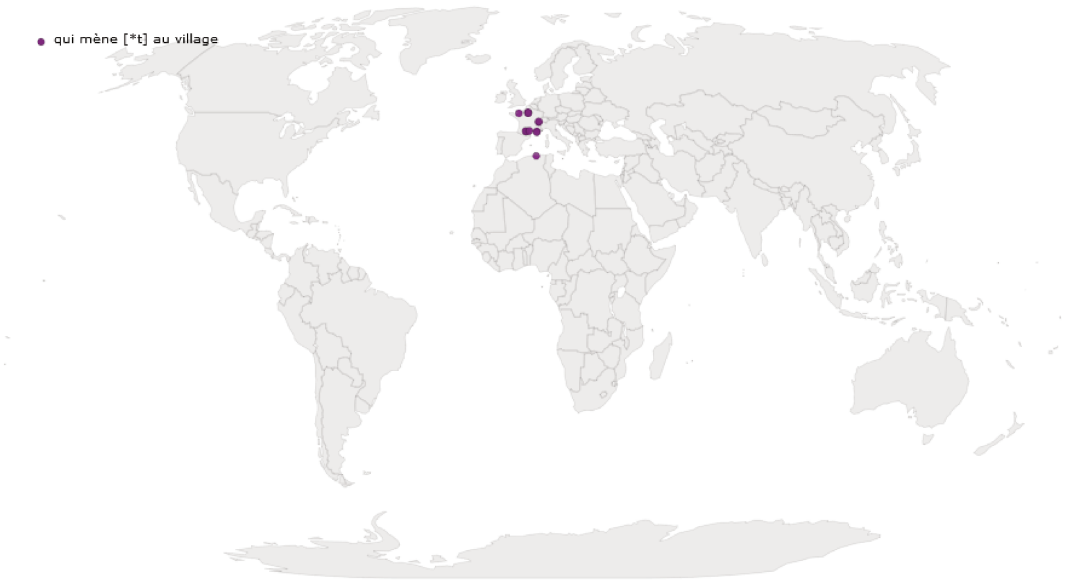


FIGURE 4.4 : Enquêtes de PFC où le pataquès *qui mène* [*t] *au village* a été produit dans la tâche de lecture du texte

(242) LL : (bruit de bouche) Je sais pas, ces temps-ci je suis14t assez facile à décevoir là, mais euh. (PFC, Trois-Rivières (Canada, Amérique du Nord), 2014, cqell1)

(243) GM1 : pis que tu es11t assis là dehors tu prends10 une petite bière tu regardes, (PFC, cqbgm1)

(244) MA1 : Inconsciemment vu que tu es11t endormie.
(PFC, Saguenay (Canada, Amérique du Nord), 2006, cqbma1)

(245) CIAAK1 : Parce que, pour faire ça aussi, il faut10 avoir un grand11nVN espace parce que le feu <E1 : Ah, bon.> qu'il y a.
(PFC, Abidjan (Côte d'Ivoire, Afrique subsaharienne), 2006, ciaak1)

(246) JG : avec un vrai14t investissement, là.
(PFC, Trois-Rivières (Canada, Amérique du Nord), 2014, cqejg1)

(247) BFASY1G : un11nVN ancien24z étudiant vient se présenter20 à toi, et puis, euh, 'moi je suis10 un tel, bon' <E : Vous avez vu la fille tout à l'heure là ?>.
(PFC, Burkina Faso (Afrique subsaharienne), 2007, cqbma1)

(248) CP : des11z anciens24n esclaves, des11z anciens21z esclaves qui se sont10 enfuis. Et euh, du coup ils11z ont fait des petites communautés là-bas quoi.

(PFC, Dijon (France, Europe), 2007, 21acp1)

(249) MG : ils rénovent les, les, les11z immeubles de bas14t en haut, de bas11z en haut et,
(PFC, Lyon (France, Europe), 2004, 69amg1)

(250) JD : Et temps11z en, d/ dès qu'il fait nuit ben tous les euh14z alligators ben ils sortent. C'est10 à ce moment là qu'ils vont pêcher donc
(PFC, Genève (Suisse, Europe), 2009, sgajd1)

(251) GM1 : Euh plus tard je leur11z avais demandé/ je leur disais toujours quand ils10 étaient petits,
(PFC, cqbgm1)

(252) RK1 : Les conseillers ne leurs14z aident pas10 à remédier à ce qu'ils11z ont dit auparavant.
(PFC, RCA Bangui (République centrafricaine, Afrique subsaharienne), 2008, rcark1)

(253) CIASB1L : Pourquoi, parce que, il y a des gens qui11z arrivent, ils veulent pas10 attendre.
(PFC, Abidjan (Côte d'Ivoire, Afrique subsaharienne), 2006, ciasb1)

Ces pataquès apparaissent dans le contexte *suis/es* + X dans les variétés canadiennes (exemples 230 à 244), dans le contexte NUM + NOM (exemples 110, 111 et 112 p.170), dans le contexte ADJ + NOM (exemples 245 à 248, les [*n] apparaissant après les M1 à voyelle nasale finale *grand* et *anciens*[15]), au sein d'une locution figée (en 249, *de bas* [*t] *en haut* est produit avant réparation), après une marque d'hésitation (*euh* [*z] *alligators* en 250), dans le contexte PRO *leur* + VER (exemples 251 et 252) et dans le contexte complémenteur *qui* + VER.3.PL (exemple 253). On note quelques surprenantes absences présumées, comme les pataquès dans le contexte (VER.PRS.IMP.2.SG) + enclitique + enclitique, présents dans le corpus ALIPE (exemples 66 p.117 et 202 p.234). Le caractère dense de ce dernier et son ancrage actionnel et situationnel (Koch & Oesterreicher, 2001 : 586) pourraient favoriser l'émission d'ordres, contrairement aux dialogues de PFC avec détachement actionnel et situationnel (Koch & Oesterreicher, 2001 : 586).

Un bilan quantitatif des occurrences de pataquès du corpus PFC est présenté dans le tableau 4.6. Tout comme pour le corpus ALIPE, on remarque que globalement, les pataquès y sont rares, puisqu'on ne les rencontre que dans 0,11% (57/53861[16]) des

[15] Notons la réparation en 248.

[16] Le nombre total de contextes de liaison et de pataquès est obtenu sur l'interface de

contextes de liaison et de pataquès[17]. À nouveau, rappelons-nous que ces résultats dépendent des pataquès retenus comme tels et que certaines données du corpus peuvent rester silencieuses. Dans ce corpus, on observe davantage de pataquès lors de la tâche de lecture du texte (0,27% ; 33/12193[18]) que dans les conversations (0,06% ; 24/41668[19]).

Lecture		**Conversations**		**TOTAL**	
%	n	%	n	%	n
0,27	33/12193	0,06	24/41668	0,11	57/53861

TABLEAU 4.6 : Bilan quantitatif de la première fouille du corpus PFC (modifié depuis Coutanson, 2020 : 10 après réanalyse)

Pour conclure, lors de cette première fouille du corpus PFC, le bruit a pu à nouveau être éliminé, sans qu'il ait été possible d'agir sur le silence. L'annotation 4, ou liaison épenthétique, m'a permis d'obtenir la plupart des occurrences de pataquès ici présentées, mais j'ai également usé d'autres requêtes afin d'obtenir quelques réalisations supplémentaires. Cette fois-ci, je n'ai pas pris le temps de m'arrêter sur les occurrences que je ne retenais pas. Notons seulement que soit je conteste l'annotation retenue (perceptiblement, aucun pataquès ne semble produit), soit ces occurrences ne sont pas des pataquès (il s'agit de schwas, de liaisons, de *quand* devant initiale consonantique…). Cette première fouille du corpus s'est elle aussi révélée fructueuse puisque des pataquès ont été mis au jour aussi bien en lecture, où ils étaient davantage produits, que dans les interactions. Comme le note Côté (2017 : 21), la plupart des exemples proviennent de variétés non européennes. Une fois de plus, cette étape s'est révélée chronophage, mais elle s'est avérée cruciale, puisque les pataquès auparavant relevés lors de l'étape de carottage étaient très spécifiques.

Lors de cette phase initiale de carottage et de première fouille des données, j'ai commencé par présenter les cinq corpus de ma sélection, puis la méthodologie que

recherche du corpus PFC en ne tenant pas compte des données incertaines et des données produites lors de la tâche de lecture de la liste de mots, à l'aide de la requête `@type ("t" | "g" | "l") @liaison_presence (1 | 0 | 3 | 4)`.

[17] voir note de bas de page 11 p.11

[18] Le nombre total de contextes de liaison et de pataquès est obtenu pour les données de la tâche de lecture du texte en ne tenant pas compte des données incertaines, à l'aide de la requête `@type "t" @liaison_presence (1 | 0 | 3 |4)`.

[19] Le nombre total de contextes de liaison et de pataquès est obtenu pour les données de la tâche de lecture du texte en ne tenant pas compte des données incertaines, à l'aide de la requête `@type ("g" | "l") @liaison_presence (1 | 0 | 3 |4)`.

j'envisageais pour cette étape. Outre de premiers exemples de pataquès, cette phase de carottage a permis d'écarter d'autres phénomènes à la marge, et j'ai dans le même temps engagé une réflexion quant à la manière de transcrire et d'annoter les phénomènes que j'étudiais à l'aide du logiciel CLAN. Lorsque le corpus le permettait, pour ALIPE et PFC, une première fouille des données a été effectuée afin de mettre au jour des exemples additionnels. L'étude de l'ALAVAL, du fonds Charles Joisten et des corpus ALIPE et PFC s'achève ici et ils ne seront pas utilisés pour une analyse plus fine des phénomènes de pataquès et de liaison en leur sein. Seul le FJD, ainsi que le corpus de tops 1 créé après cette phase de carottage et d'annotation, seront systématiquement décortiqués dans le chapitre suivant. Nous commencerons par mentionner le cas de dénasalisations surprenantes qui apparaissent en marge du phénomène de liaison dans le FJD.

Chapitre 5

Résultats et discussion : fouille du FJD et du corpus de tops 1

5.1 Dénasalisations surprenantes liées au phénomène de liaison dans le FJD

5.1.1 Dénasalisation de la voyelle finale du M1 hors contexte de liaison

Nous avions mentionné en 1.2.2.1.4 le phénomène de dénasalisation qui peut se produire avant une liaison en [n]. Dans les exemples suivants, nous faisons face à des réalisations inattendues : un mot qui liaisonnerait en [n] en contexte de liaison est dénasalisé hors contexte de liaison, devant un mot débutant par la consonne [n]. En 254 par exemple, la voyelle de *son* est dénasalisée devant *noble*, comme si ce dernier mot était à initiale vocalique et qu'il était liaisonné. Il en va de même pour *un* et *aucun* en 255 et 256, qui voient leur (dernière) voyelle être dénasalisée devant *naufrage*.

(254) *SGR : par so(n) noble chant.
%pho : so[1] noblə ʃɑ̃
(JD_a061_04, 43810, Virginie Granouillet)

(255) *SGR : j'ai grand peur pour u(n) naufrage la tempête ou le vent.
%pho : ʒɛ gʁɑ̃ pœʁ puʁ y[2] nofʁaʒə la tɑ̃pɛtə u lə vɑ̃
[...]

[1] Dans la variété des chanteurs du FJD, la voyelle nasale [õ] semble principalement dénasalisée en [o], avec fermeture de la voyelle donc.

[2] La dénasalisation de la voyelle de *un* en [y] peut surprendre. Nous aborderons cette particularité qu'on observe dans le FJD en 5.1.2.

*SGR : j'arriv(e)rai il arriv(e)ra aucu(n) naufrage lorsque je serai sur l'eau.
%pho : ʒaʁivʁɛ ilaʁivʁa oky nofʁaʒə lɔʁskə ʒə səʁɛ syʁ lo
(JD_a064_04, 43800, Rosalie Ferret)

(256) *SGR : j'ai grand peur de les naufrages la tempête ou le vent.
[...]
SGR : n'arriv(e)ra aucun naufrage quand [@] je serai sur l'eau.
%pho : naʁivʁa ɔky nɔfʁaʒə kɑ̃tə ʒə səʁɛ syʁ lo
(JD_a161_02, 43800, Jacques Rogues)

Quelques exemples de dénasalisations de voyelles finales quand le mot suivant commence par un [n] peuvent donc être observés dans le FJD, ce qui pourrait correspondre à des métanalyses de la part des chanteurs, ces mots étant pris pour des mots à initiale vocalique liaisonnés. Nous poursuivrons avec un deuxième phénomène de dénasalisation atypique de voyelles dans le FJD, en marge du phénomène de liaison, couplé cette fois-ci d'une fermeture de la voyelle.

5.1.2 Fermeture et dénasalisation de la voyelle de *un* en contexte de liaison

Nous avions vu en 1.2.2.1.4 que l'opposition entre /œ̃/ et /ɛ̃/ avait cessé d'exister en français standard contemporain, le premier phonème disparaissant au profit du second. Dans le FJD, cette opposition s'observe encore. Dans ce corpus se laisse parfois apercevoir une prononciation particulière du déterminant indéfini masculin *un* en liaison, prononcé [œ̃] devant un mot à initiale consonantique (exemple 257) et [yn] devant initiale vocalique (exemple 258), de la même façon que le déterminant indéfini féminin *une* devant initiale vocalique lorsque le schwa n'est pas réalisé (exemple 259)[3].

(257) *SGR : y a [*t] un bal annoncé.
%pho : ja tœ̃ bal anɔ̃se
(JD_a061_14, 43810, Virginie Granouillet)

(258) *SGR : au jardin de mon pèr(e) il y a [*t] un oranger.
%pho : o jardɛ̃ də mɔ̃ pɛr il ja tyn orɑ̃ʒe
(JD_a063_23, 43810, Virginie Granouillet)

[3] Les trois exemples ici sélectionnés relèvent de la même chanteuse, Virginie Granouillet, et sont produits dans le même contexte lexical et syntaxique.

(259) *SGR : là + haut là + haut dedans la tour y a [*t] un(e) princess(e) à mes amours.
%pho : laolao dədɑ̃ la tur ja tyn prɛ̃sɛz a me zamuʁ
(JD_a063_02, 43810, Virginie Granouillet)

J'ai sélectionné tous les mots du FJD finissant par ‹un› en contexte de liaison, et je n'ai conservé que les occurrences pour lesquelles la liaison était effective, avant de procéder à une annotation de la voyelle sur une base perceptive. Dans le fonds, le phénomène observé ne se limite pas au déterminant *un*, puisqu'un autre mot, *aucun*, est concerné. Dans ces conditions, la prononciation [yn] y est importante en liaison (57,7% ; 90/157)[4].

Nous avions aussi vu en 1.2.2.1.4 que lors de la réalisation d'une liaison en [n], la voyelle nasale finale du M1 pouvait être dénasalisée : /ɛ̃/ en /ɛ/ voire /i/ et /ɔ̃/ en /ɔ/. La seule mention d'une dénasalisation de /œ̃/ en /y/ que j'ai rencontrée se trouve dans Hornsby (2020 : 85). Notons bien qu'ici, il se produit à la fois une dénasalisation et une fermeture de la voyelle. Par ailleurs, comme indiqué en 1.2.2.2.4, le genre du mot liaisonnant peut parfois être neutralisé, la forme adoptée en liaison pouvant être similaire à la forme du masculin (pour les déterminants unipossessifs) ou à la forme du féminin (pour le déterminant démonstratif singulier et quelques adjectifs prénominaux). Ici, nous sommes dans le deuxième cas, le déterminant indéfini masculin singulier *un* ayant une prononciation en liaison équivalente au déterminant indéfini féminin singulier *une*. On doute parfois du genre des mots à initiale vocalique (voir dans ce manuscrit les réalisations non normatives *une en-tête*, p.135, et *une astérisque*, p.137), car dans de nombreux contextes, il y a neutralisation au niveau du genre, que ce soit au niveau du déterminant défini élidé *l'*, des déterminants unipossessifs (*mon*, *ton*, *son*, *mes*, *tes*, *ses*), des déterminants démonstratifs (*cet*, *ces*), de la préposition *de* élidée (*d'*) ou encore de certains adjectifs prévocaliques à l'oral (*bon*, *vieil*, *nouvel*, *bel*, *petit*...). Une neutralisation au niveau du déterminant indéfini, comme ici, constitue encore un indice en moins pour orthographier correctement les mots en question (par exemple, *un ombrage* et non *une ombrage*), mais également à l'oral afin d'associer au nom en question certains mots dont le genre est audible (comme des adjectifs épithètes) ou de réaliser une reprise à l'aide d'un pronom anaphorique.

Pour conclure, les chanteurs du FJD présentent la spécificité de fermer et de dénasaliser la voyelle d'*un* ou la voyelle finale de mots graphiques finissant en ‹un› lorsqu'une

[4] Je ne me suis pas intéressée à la nasalisation de la voyelle /œ/. J'ai également relevé une occurrence de [ɔ̃] : il s'agissait d'une harmonie vocalique.

liaison est réalisée, entraînant la neutralisation du M1. Pour dépasser l'annotation sur une base perceptive, on pourrait favoriser une analyse acoustique des données. Sur un spectrogramme, on pourrait notamment analyser le F1 des voyelles, celui-ci étant habituellement plus bas pour /y/ que pour /œ/ (Georgeton et al., 2012 : 149)[5]. Notons tout de même qu'une étude acoustique de 2018 comparant les voyelles de l'anglais américain dans la voix parlée et la voix chantée[6] a montré que F1 était, pour cette langue, plus bas en chant que dans la parole (Bradley, 2018 : 387). La différence était significative uniquement chez les femmes, qui se trouvaient être les interprètes les plus formées en technique vocale de cette étude (Bradley, 2018 : 387). De la sorte, des /œ/ ou /œ̃/ pourraient avoir tendance à tendre vers /y/ ou /ỹ/ en chanson, mais si tel était le cas, des /yn/ perçus devant initiale vocalique devraient de la même façon être perçus /ỹ/ devant initiale consonantique.

Après avoir fait la rencontre de deux phénomènes de dénasalisation surprenants dans le FJD, nous allons regarder de plus près quelques phénomènes erratiques que présentent le FJD et le corpus de tops 1, en commençant par quelques non-réalisations inattendues de liaisons.

5.2 Productions erratiques en matière de liaisons dans le FJD et le corpus de tops 1

5.2.1 Non-liaisons attestées surprenantes dans le FJD et le corpus de tops 1

Dans les deux corpus, on note ça et là quelques non-réalisations surprenantes de liaisons, dans des contextes qui seraient aujourd'hui considérés comme invariables. Dans la littérature, de telles réalisations sont parfois considérées comme des *erreurs de performance* (cf par exemple Morin & Kaye, 1982 : 295). La notion de *performance* nous vient de la linguistique générative, et s'oppose à la notion de *compétence* (Chomsky, 1965 : 4). Cette opposition présente un lien de parenté avec l'opposition saussurienne entre *langue* et *parole* (Saussure, 1916). La **compétence** est définie par Chomsky comme "the speaker-hearer's knowledge of his language" et la **performance** comme "the actual use

[5]Indiquons cependant que l'étude de Georgeton et al. (2012) porte uniquement sur les voyelles orales du français. Or, en liaison, les autres réalisations d'*un*, excepté /yn/, sont habituellement nasalisées.

[6]Cette étude présente un intérêt tout particulier ici, car le matériel utilisé pour éliciter la voix chantée est une chanson folklorique.

of language in concrete situations" (Chomsky, 1965 : 4). Les générativistes mettent l'accent tout particulièrement sur la compétence, en considérant la performance comme relevant de l'individuel et de l'hétérogène. En effet, bien que Chomsky (1965 : 4) admette que "[a] record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on", selon lui "[t]he problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance". Ainsi, dans ce cadre théorique, les occurrences en question sont agrammaticales. Il s'agit d'erreurs de production qui ne devraient pas être réitérées. Au contraire, dans une linguistique fondée sur l'usage, les réalisations d'apparence marginales trouvent leur place au sein du système qu'est la langue. Notamment, une approche constructionniste "does not assume that language should be divided up into 'core' grammar and the to-be-ignored 'periphery.'" (Goldberg, 2003). Il ne s'agit alors pas d'erreurs, mais de réalisations erratiques qui pourraient s'expliquer en observant la langue dans sa globalité.

Le corpus de tops 1 tout d'abord présente une occurrence de disjonction qui n'aurait plus cours, il me semble, de nos jours : le contexte *mon* + *igloo* n'est en effet pas liaisonné par le chanteur Plastic Bertrand (exemple 260). De telles disjonctions se laissent apercevoir également dans le FJD, comme dans *mes* [oz] *entrailles* (exemple 261), *vous* [oz] *entendrez* (exemple 262) et *on* [on] *en use pas* (exemple 263)[7].

(260) *SGR : comme un indien dans son [oz] igloo.
%pho : kɔm ɛ̃ nɛ̃djɛ̃ dɑ̃ sɔ̃ iglu
(Tops 1, Plastic Bertrand, *Ça plane pour moi*, 1978 [1977])

(261) *SGR : tou(s) mes [oz] entrailles et mes boyaux.
%pho : tu me ɑ̃tʁaːj e me bɔjo
(JD_a359_03, 43800, Mme Rostin)

(262) *SGR : vous [oz] entendrez parler de vous.
%pho : vu ɑ̃tɑ̃dre parle də vu
(JD_a163_13, 43800, Jean-Pierre Arnaud)

(263) *SGR : point de chandell(e) i(l) n'en faut pas dans notr(e) couvent on [on] en use pas.

[7]Dans certains cas, on remarque la présence d'un [h] en contexte de liaison. Ces réalisations n'ont pas été annotées et ont été considérées comme des non-liaisons. Pourtant, nous avions vu en 1.2.3.3 que dans la variété louisianaise de la paroisse Lafourche, les [h] pouvaient correspondre à une évolution de la consonne de liaison [z].

%pho : pwɛ̃ də ʃɑ̃dɛl il ɑ̃ fo pa dɑ̃ nɔtʀ kuvɑ̃ hɔ̃ ɑ̃ nyzə pa
(JD_a165_36, 43800, Valentine Alibert)

Nous avons pu observer quelques non-réalisations inattendues de liaisons dans le corpus de tops 1 et le FJD. Notre attention se portera désormais sur quelques liaisons réalisées en contexte disjonctif dans ces deux corpus.

5.2.2 Liaisons en contexte disjonctif dans le FJD

Nous étudierons ici quelques cas parfois nommés *psiloses* (pour une critique du terme, voir 1.2.3.1.1.3). Dans le FJD, les seules occurrences des mots *Hollande* (264), *Hollandais* (265) et *Hongrie* (266) sont liaisonnées[8]. On pourrait se dire qu'il s'agit de toponymes et gentilés éloignés de la réalité des locuteurs, et qu'ils ne savent pas toujours comment liaisonner ces mots, qui sont usuellement disjonctifs. En effet, d'autres toponymes lointains ont des prononciations non standard dans le fonds, comme *Algérie/Alger* prononcés [alɛʒɛʀə]/[aləʒe] (JD_a168_11), *Égypte* prononcé [eʒipə] (JD_a060_04) ou encore Bethléem, prononcé [betɛlɛɛm] (JD_a043_16). Le caractère aujourd'hui disjonctif de *Hongrie*, malgré son origine latine (voir 1.2.3.1.1.3), est frappant, par rapport à *Hollande* qui vient lui bien d'une langue germanique (néerlandais). Il est donc permis de douter que la disjonction ait toujours été faite avec *Hongrie*. Une rapide recherche dans Google Books[9] montre que ces mots ont pu être graphiés sans ‹h› initial ; on note aussi quelques élisions.

(264) *SGR : oh@i j'irais en [n] Hollande.
(JD_a165_04, 43800, Germaine Varenne)

(265) *SGR : pris par les [z] Hollandais qu'ell(e) me verra jamais.
(JD_a160_11, 43800, Jean-Pierre Arnaud)

(266) *SGR : en Franc(e) en Itali(e) en Savoi(e) en [n] Hongri(e).
(JD_a066_17, 43800, Rosalie Ferret)

Un autre mot à ‹h› initial est toujours liaisonné en contexte de liaison dans le FJD : il s'agit du nom propre *Henriette*, d'origine germanique, qu'on retrouve dans les différentes versions d'une même chanson interprétée par plusieurs chanteurs (exemples 267, 268 et 269), ainsi que dans une chanson non apparentée (exemple 270). On peut noter

[8] *Hollandaise*, *Hollandaises*, *Hongrois*, *Hongroise* et *Hongroises* étaient absents du fonds.
[9] Google Books est accessible à l'adresse suivante : https ://books.google.com/.

qu'une de ces liaisons est un pataquès (*y conduit* [*z] *Henriette* en 268). Ces liaisons apparaissent dans le même contexte morphosyntaxique (VER.PRS.3.SG + *Henriette*) et, à l'exception du pataquès, impliquent une même consonne, [t]. Dans la littérature, Fouché (1956 : 457-458) indique pourtant que les liaisons ne se font pas « sauf les exceptions indiquées dans les remarques ci-dessous entre une forme verbale [...] et un mot suivant ». La pléthore d'exemples que cet auteur se plaît à présenter permet d'en repérer un similaire à mes observations : *Que dit | Henri ?* (Fouché, 1956 : 458). Néanmoins, il présente l'exception suivante nous concernant : « [o]n fait cependant la liaison entre *c'est, c'était, ce serait* et un nom propre » (Fouché, 1956 : 460). Sur ce point précis, les propositions de Fouché ne se révèlent donc pas valables pour le FJD, puisque sa règle et ses exceptions n'autoriseraient que l'exemple 270.

(267) *SGR : jeunesse trop coquette écoutez la leçon que vous fait [t] Henriette et son amant Damon.
[...]
*SGR : un secret mouvement y conduit [t] Henriette.
(JD_a063_20, 43810, Virginie Granouillet)

(268) *SGR : que vous fait [t] Henriett(e) et son amant Damon.
[...]
*SGR : un secret mouvement y conduit [*z] Henriette.
(JD_a287_12, 63550, Marie Fontenille)

(269) *SGR : que nous dit [t] Henriette et son amant Damon.
(JD_a163_10, 43800, Jean-Pierre Arnaud)

(270) *SGR : buvons don(c) la plu(s) jeun(e) c'est [t] Henriette.
*SGR : la plu(s) jeun(e) c'est [t] Henriette.
(JD_a154_05, 43800, André-Joseph Aguilhon)

J'ai ici relevé quelques exemples dans le FJD à ‹h› initial pour lesquels on attendrait aujourd'hui une disjonction, questionnant nos certitudes concernant les M2 pouvant être disjonctifs, en synchronie comme en diachronie. D'autres réalisations surprenantes de liaisons seront mises en avant dans le point suivant, dans des contextes où des liaisons ne sont habituellement pas attestées.

5.2.3 Liaisons attestées surprenantes dans le FJD et le corpus de tops 1

Dans le FJD, les locutions prépositives *par/pour rapport (à)* peuvent contenir une liaison (les exemples 272, 273, 275, 276 et 277 sont liaisonnés et les exemples 271 et 274 ne le sont pas ; cette collection d'exemples est exhaustive). Pourtant, Fouché (1956 : 446) indiquait que la liaison ne se faisait pas dans *par rapport à*. Le FJD révèle ainsi des liaisons surprenantes, dont l'existence n'a jamais été envisagée, d'où la nécessité, comme proposé en 2.1.3.4.2, d'avoir une définition large du contexte de liaison afin d'envisager des liaisons inattendues en fonction de nos connaissances actuelles du phénomène et de permettre une meilleure comparabilité inter-corpus.

(271) *SGR : c'est mon cruel de père c'est ma cruell(e) de mère.
*SGR : par rapport [ot] à vous.
*SGR : par rapport [ot] à vous.
(JD_a043_21, 43810, Marie Soulier)

(272) *SGR : mon cruel de pèr(e) et ma cruell(e) de mère par rapport [t] à vous.
*SGR : par rapport [t] à vous.
(JD_a060_19, 43810, Virginie Granouillet)

(273) *SGR : c'est mon cruel de père qui m'a renfermée par rapport [t] à vous.
(JD_a156_03, 43800, Fanny Maligue)

(274) [...] *SGR : qui pourra vous loger par rapport [ot] à son fils.
(JD_a165_24, 43800, Germaine Varenne)

(275) *SGR : c'est mon très cher pèr(e) qui m'a renfermé(e) par rapport [t] à[10] vous.
(JD_a347_08, 43200, Séraphine Chaize)

(276) *SGR : leur a dit montez messieurs là + haut sur c(e) beau château v(ou)s y trouverez.
*SGR : la mèr(e) de mon mari qui pourra vous loger pour rapport [t] à son fils.
(JD_a355_01, 43800, Marguerite Rousset)

(277) *SGR : par rapport [t] un amant jaloux.
(JD_a359_01, 43800, Mme Rostin)

En ce qui concerne le contexte NOM singulier + , Delattre (1947 : 152) y considérait les liaisons comme « interdites ». Il indiquait toutefois que la liaison était « obligatoire » dans le « groupe figé » *nuit et jour* (Delattre, 1947 : 156-157). Toutes les occurrences de cette locution adverbiale sont bien liaisonnées dans le FJD (six occurrences). Mais

[10] Cet exemple n'apparaît pas dans le corpus puisque ce vers est déclamé et non chanté.

la liaison après *nuit* ne se limite pas à ce contexte dans le fonds, puisque *la nuit et le jour* peut également être liaisonnant (278, 279, 280 et 282 sont liaisonnés, 281 ne l'est pas ; cette collection d'exemples est exhaustive)[11]. Les liaisons après un nom singulier impliquant la consonne [n] avaient pour leur part déjà été mentionnées lors du carottage du FJD (*le matin* [n] *et le soir* dans l'exemple 157 p.220). L'exemple 283 tient lieu d'exemple supplémentaire[12]. On remarquera aussi la liaison en [z] après le nom singulier *flux* (exemple 284) et de façon plus surprenante celle en [r] après le nom singulier *monsieur* (285).

(278) *SGR : les garçons de chez nous grand Dieu qu'i(l)s ont de peine la nuit [t] et le jour.
(JD_a061_12, 43810, Virginie Granouillet)

(279) *SGR : grand Dieu qu'ils ont de peines la nuit [t] et le jour.
(JD_a205_11, 43810, Marie Soulier)

(280) *SGR : les garçons de chez nous grand Dieu qu'ils ont de peines la nuit [t] et le jour.
(JD_a329_02, 43350, Joseph Bonnefoux)

(281) *SGR : travaillant la nuit [ot] et le jour.
(JD_a075_23, 43810, Marie Soulier)

(282) *SGR : travaillait la nuit [t] et le jour.
(JD_a205_14, 43810, Marie Soulier)

(283) *SGR : hélas que tu es folle disait le vin [n] à l'eau.
(JD_a066_17, 43800, Rosalie Ferret)

(284) *SGR : et mon.
*SGR : flux flux [z] et reflux.
(JD_a066_17, 43800, Rosalie Ferret)

(285) *SGR : le monsieur [r] et la dame.
(JD_a168_10, 63550, Marie Rigodias)

Un cas similaire peut être noté dans le corpus de tops 1, toujours dans le contexte NOM singulier + X. S'il est bien connu que la locution adverbiale *de temps en temps* est liaisonnante, le chanteur Johnny Hallyday produit une occurrence inattendue entre

[11] Notons que 278, 279 et 280 correspondent à trois versions d'une même chanson, quand 281 et 282 sont deux performances d'une même chanson par une même chanteuse.

[12] Les exemples préverbaux, par précaution, ne sont pas présentés.

temps et *encore* (286). Fouché (1956 : 440-448) indiquait cependant que la liaison était faite dans *il est temps* [z] *encore*, mais aussi dans *temps* [z] *affreux*, *en temps* [z] *et lieu*, *quelque temps* [z] *après*, *quelque temps* [z] *auparavant*, *de temps* [z] *à autre* et *le temps* [z] *et l'espace*. Dans le corpus PFC, seules les locutions *de temps en temps* et *de temps à autre* présentent une liaison. Cette réalisation de Johnny Hallyday nous montre à nouveau que malgré leur caractère figé, les locutions semblent pouvoir conserver une certaine influence analogique.

(286) *SGR : combien de temps [z] encore ?
(Tops 1, Johnny Hallyday, *Oh ! Ma jolie Sarah*, 1971 [1971])

Parmi les contextes considérés comme non liaisonnants dans la littérature, d'autres révisions restent à faire. Delattre (1947 : 156) décrivait par exemple les liaisons dans le contexte ADJ singulier + invariable comme « interdites », ce qu'il illustrait avec les exemples suivants : *brun et gris* et *sot et méchant*. *Doux* [z] *et malin* est pourtant liaisonné dans le FJD (exemple 287). Le même cas de figure s'observe à nouveau pour le contexte PRO nasal + invariable, dans lequel selon cet auteur les liaisons sont « interdites », à l'image de l'exemple suivant qu'il mettait en avant : *va-t'en à ta place* (Delattre, 1947 : 156). Or, comme on le voit dans les exemples 288 et 289, cette affirmation de Delattre n'est pas confirmée pour le FJD, puisque ce contexte peut y être liaisonnant. De la même façon, selon Fouché (1956 : 462), la liaison ne se fait pas « [e]ntre [l'adverbe] [...] DEMAIN [...] et un mot suivant ». Le FJD présente néanmoins un contre-exemple : *demain* [n] *à l'aurore* (exemple 290). Par ailleurs, selon Fouché (1956 : 442-443), la liaison n'est pas faite « entre un substantif singulier ou pluriel [...] et les pronoms adverbiaux *en*, *y* », affirmation illustrée à l'aide des exemples suivants : *Les gens | y sont allés* ou *Les journaux | en parlent*, une fois de plus infirmée par les données du FJD (exemples 291 et 292). Un dernier cas d'étude : selon Fouché (1956 : 443), les liaisons peuvent être réalisées dans le contexte NOM pluriel + VER lorsqu'on adopte un style soutenu, quand Delattre (1947 : 153) estimait que les liaisons y sont « facultatives ». Nous avions vu en 1.2.2.3.2 que dans le corpus PFC, les réalisations dans ce contexte n'étaient pas attestées. Il n'en va pas de même dans le FJD, où on peut relever des exemples tels qu'en 293[13].

[13] Ici, on pourrait penser qu'il s'agit du gentilé *Françaises* au lieu de *Français*. Ceci est vite balayé puisqu'il est ici question de militaires. Nous rencontrons à nouveau un cas de neutralisation du genre au niveau du M1, le gentilé masculin *Français* étant prononcé en liaison de la même façon que le gentilé féminin *Françaises*. On pourrait encore plaider en faveur de la présence d'un pronom *ils* dont la seule trace visible serait une liaison en [z] : *que les Français (il)s arrivent*.

(287) *SGR : d'un air doux [z] et malin puis il [*@] s'envole.
(JD_a329_04, 43350, Léonie Bonnefoux)

(288) *SGR : médecin va t'en [n] au diable moi que j'aime tant le vin.
*SGR : médecin va t'en [n] au diable oh@si !
(JD_a330_04, 43800, Henriette Usson)

(289) *SGR : va t'en [n] au port des armes !
*SGR : va t'en [on] au port des armes dans ton beau régiment !
(JD_a330_05, 43800, Henriette Usson)

(290) *SGR : tu lui feras mes compliments demain [n] à l'aurore.
(JD_a061_04, 43810, Virginie Granouillet)

(291) *SGR : car mes pieds [z] en sont remplis de sang.
(JD_a064_07, 43800, Rosalie Ferret)

(292) *SGR : les amants [z] y vont.
(JD_a328_02, 43800, André-Joseph Aguilhon)

(293) *SGR : c'est pour faire savoir que les Français [z] arrivent.
(JD_a066_07, 43800, Rosalie Ferret)

Nos connaissances actuelles du phénomène de liaison ont à nouveau été remises en cause par plusieurs réalisations du FJD et du corpus de tops 1, par exemple dans les locutions prépositives *par/pour rapport (à)* et dans les contextes NOM.SG + X, ADJ.SG + KON, PRO *en* + PRP, *demain* + X et NOM.PL + ADV *en/y*, qui ne sont habituellement pas considérés comme liaisonnants. J'ai enfin montré que contrairement à PFC, le FJD présentait des réalisations dans le contexte NOM.PL + VER. Il ne s'agit là que de quelques observations, et les données pourraient encore apporter de nouvelles attestations surprenantes. En décelant une liaison en dehors de la locution adverbiale de *temps en temps*, je me suis aussi interrogée quant à la possibilité d'une influence analogique des locutions, malgré leur caractère figé. Finalement, j'ai rappelé une fois de plus la nécessité d'annoter tous les contextes de liaison des corpus, même si jusque-là aucune liaison n'a été relevée dans ces contextes, pour une bonne comparabilité inter-corpus.

Dans cette partie, nous nous sommes intéressés à trois types de productions erratiques en matière de liaison dans le FJD et le corpus de tops 1 : des non-réalisations en contexte invariable, des réalisations devant des mots disjonctifs et des réalisations dans des contextes dans lesquels les liaisons ne sont habituellement pas attestées. Dans la

partie suivante, nous nous pencherons sur le lexique employé comme M1, en regardant tout d'abord si nos corpus présentent des adjectifs masculins prénominaux n'ayant pas été observés jusque là en contexte de liaison.

5.3 Liaison et lexique : actualisation des connaissances à l'aide des données du FJD et du corpus de tops 1

5.3.1 Adjectifs masculins prénominaux en contexte de liaison dans le FJD et le corpus de tops 1

En 1.2.2.2.2, nous avions vu que dans sa thèse de doctorat, Barreca (2015 : 191) n'avait relevé que douze adjectifs masculins prénominaux en contexte de liaison dans les grands corpus PFC et CFPP2000 : *ancien*, *bas*, *bon*, *certain*, *grand*[14], *gros*, *mauvais*, *moyen*, *petit*, *plein*, *prochain* et *saint*. Mes corpus sont de taille bien plus modeste, et je n'observe que sept adjectifs masculin prénominaux dans le corpus de tops 1 : *bel* [15], *dernier*[16], *grand*, *nouvel*[17], *parfait*, *petit* et *vieil*[18] et huit dans le FJD : *bel*[19], *bon*, *charmant*, *grand*, *maudit*, *odieux*, *parfait* et *petit*. Ainsi, les adjectifs supplémentaires à apparaître en position prévocalique seraient *bel*, *charmant*, *dernier*, *nouvel*, *maudit*, *odieux*, *parfait* et *vieil*. Pourtant, une fouille rapide de PFC montre que *dernier* (*dernier* [ʁ] *étage* ; Lyon (France, Europe), 2004, 69amg1) et *parfait* (*parfait* [ot] *allemand*, *parfait* [ot] *anglais* : Neuchâtel (Suisse, Europe), 2011, scapm1) y sont bien présents et que ces données étaient disponibles lorsque Barreca rédigeait sa thèse de doctorat. On relève aussi dans ce corpus les adjectifs prénominaux *bel* (*bel* [l] *intérieur* ; Gembloux (Belgique, Europe), 2007, bgadb1), *nouvel* (*nouvel* [l] *environnement* ; Trois-Rivières (Canada, Amérique du Nord), 2014, cqecc1) et *vieil* (*vieil* [j] *ordi* ; Trois-Rivières (Canada, Amérique du Nord), 2014, cqead1). Ces adjectifs n'étant habituellement pas annotés pour la liaison, l'oubli est compréhensible. Enfin, j'ai un doute concernant les adjectifs *saint* et *moyen*. Le premier semble apparaître uniquement au sein de toponymes dans les deux corpus, et le second exclusivement dans la locution nominale *Moyen-Âge*, ce qui n'en fait pas selon moi des éléments comparables au reste des adjectifs prénominaux relevés. Concernant les adjec-

[14] en cas de pataquès

[15] *bel* [l] *habit*

[16] *dernier* [or] *whisky*

[17] *nouvel* [l] *espoir*

[18] *vieil* [j] *air*, *vieil* [j] *amant* et *vieil* [j] *habitué*

[19] *bel* [l] *âge*, *bel* [l] *agneau*, *bel* [l] *ami*, *bel* [l] *anneau*, *bel* [l] *enfant*, *bel* [l] *officier* sans compter *bel* [l] *et bien*

tifs prénominaux *parfait* (exemples 294, 295, 296[20] et 297), *charmant* (exemple 298) et *maudit*[21], ils lient dans le FJD et le corpus de tops 1, mais les exemples sont erratiques et on ne peut pas généraliser. *Odieux*, pour sa part, ne lie pas dans l'exemple 299 issu du FJD.

(294) *SGR : et d'un parfait [t] amour ell(e) était le modèle.
(JD_a063_20, 43810, Virginie Granouillet)

(295) *SGR : et d'un parfait [t] amour ell(e) était le modèle.
(JD_a163_10, 43800, Jean-Pierre Arnaud)

(296) *SGR : et du parfait [t] amour ils étaient le modèle.
(JD_a287_12, 63550, Marie Fontenille)

(297) *SGR : vous deu(x) nouez filiez mon parfait [t] amour.
(Tops 1, Patrick Fiori, Jean-Jacques Goldman et Christine Ricol, *4 mots sur un piano*, 2007 [2007])

(298) *SGR : d'un vrai guerrier vous avez les attraits d'un charmant [t] officier.
(JD_a075_20, 43810, Marie Soulier)

(299) *SGR : de l'odieux [?] assassinat.
(JD_a061_07, 43810, Virginie Granouillet)

Pour résumer, en repartant des observations de Barreca (2015 : 191) et en y apportant quelques corrections, on peut dire que les corpus PFC et CFPP2000 ne présentent que quinze adjectifs masculins prénominaux : *ancien*, *bas*, *bel*, *bon*, *certain*, *dernier*, *grand*, *gros*, *mauvais*, *nouvel*, *parfait*, *petit*, *plein*, *prochain* et *vieil*, auxquels on peut ajouter *moyen* et *saint* en prenant en compte les locutions nominales et les toponymes. Dans le FJD, *charmant*, *maudit* et *odieux* peuvent aussi être notés. Dans le FJD et le corpus de tops 1, *charmant* (*charmant officier*), *maudit* (*maudit Anglais*) et *parfait* (*parfait amour*) lient. Les données de mes deux corpus chantés ne semblent pas assez massives pour observer davantage d'adjectifs masculins prénominaux, sans oublier que le FJD comporte de nombreux doublons. À présent, nous allons actualiser les listes de M1 qu'on observe en corpus pour les consonnes de liaison les moins fréquentes, /p/, //ʁ// et /n/.

[20] Les exemples 294, 295 et 296 proviennent du même vers, extrait de trois versions d'une même chanson, interprétée par trois chanteurs différents.

[21] On relève six occurrences de *maudit* [t] *Anglais*, réalisées par deux chanteuses interprétant une même chanson.

5.3.2 Mots 1 liaisonnants pour les consonnes de liaison les moins fréquentes dans le FJD et le corpus de tops 1

Nous avions vu en 1.2.2.4 que les liaisons les moins fréquentes étaient les plus liées à des lexèmes particuliers. Par exemple, /p/ ne s'observait en liaison dans PFC qu'après les adverbes *trop* et *beaucoup*. Seul *trop* se laisse observer dans le FJD (*trop attendu*) et le corpus de tops 1 (*trop aimé*). Notons que dans ce point, je ne présente ni les données incertaines, ni les sandhis externes préconsonantiques.

Les données de //ʁ// sont plus intéressantes : dans PFC seuls les lexèmes *se débrouiller*, *premier* et *dernier* étaient liants en conversation (voir 1.2.2.4). Dans le corpus de tops 1, j'ai relevé des liaisons seulement après des verbes. Les lexèmes représentés sont nombreux : *aller*, *s'en aller*, *se brûler*, *se cacher*, *se câliner*, *chanter*, *chercher*, *confier*, *danser*, *donner*, *échapper*, *épouser*, *frapper*, *gueuler*, *jouer*, *oser*, *passer*, *penser*, *se réchauffer*, *rester*, *rigoler* et *songer*. Dans les données du FJD, des liaisons sont également produites après de nombreux infinitifs : *accorder*, *aller*, *s'attacher*, *chanter*, *coucher*, *danser*, *déposer*, *donner*, *se fier*, *frapper*, *s'habiller*, *loger*, *louer*, *mener*, *penser*, *rester*, *sonner*, *taper* et *trouver*, après des participes passés : *aimé*[22], *allés*[22], *blessée*[22], *brodé*[22], *conduit*[22], *couché(s)*[22], *dansé*[22], *demandé*[22], *habillé(e)s*[22], *levés*[22], *logé*[22], *portée*[22] et *rencontré*[22], après un verbe à l'imparfait : *chantait*[22], devant des participes passés : *animé(e)*[22], après un adjectif : *premier* et après des noms : *mariée*[22] et *monsieur*[23]. En regroupant les infinitifs, les participes passés et les verbes à l'imparfait après lesquels des liaisons ou des pataquès peuvent être produits dans le FJD, on obtient les lexèmes *accorder*, *aimer*, *aller*, *s'attacher*, *blesser*, *broder*, *chanter*, *conduire*, *(se) coucher*, *danser*, *demander*, *déposer*, *donner*, *se fier*, *frapper*, *s'habiller*, *lever*, *loger*, *louer*, *mener*, *penser*, *porter*, *rester*, *rencontrer*, *sonner*, *taper* et *trouver*. Ainsi, on note surtout dans ces deux corpus chantés que les liaisons en //ʁ// sont produites après de multiples verbes à l'infinitif, contrairement au français parlé. Quant aux pataquès en [*ʁ] du FJD, nombreux sont ceux qui sont produits après des participes passés (cela avait été abordé en 1.3.5.1.3).

En ce qui concerne les liaisons en /n/, elles n'apparaissent en conversation dans le corpus PFC qu'après les lexèmes *ancien*, *aucun*, *b(i)en*, *bon*, *certain*, *en*, *grand*, *mon*, *non*, *on*, *plein*, *rien*, *son*, *ton*, *un* (et après les nombres finissant en *un*, voir 1.2.2.4). Le corpus de tops 1 ne fait pas preuve d'originalité, puisque j'y relève à nouveau les lexèmes *aucun*, *bien*, *en*, *mon*, *on*, *rien*, *son*, *ton* et *un* en liaison. Dans le FJD, ce type de liaisons

[22] en cas de pataquès

[23] Certaines occurrences de [*r] n'ont pas été conservées, par sécurité, comme des pataquès : il s'agirait alors d'autres sandhis externes consonantiques.

peut être noté après les lexèmes *amant*[22], *aucun*, *bien*, *bon*, *demain*, *en*, *et*[22], *matin*, *mon*, *moutons*[22], *on*, *partons*[22], *rien*, *son*, *sont*[22], *ton*, *un*, *vin* et *vrai*[22]. Les liaisons après *demain*, *matin* et *vin*, qui ont été mentionnés en 5.2.3, sont caractéristiques du FJD. Quant aux pataquès en [*n] du fonds, ils se terminent souvent par une voyelle nasale (*amant*, *moutons*, *partons* et *sont*), qui a pu influencer leur réalisation, à la manière des liaisons en [n] qui apparaissent après des voyelles nasales.

Le lexique employé comme M1 dans ces corpus chantés, dans les cas où il y a liaison, présente donc une particularité notamment pour la consonne //ʁ//. En effet, contrairement aux grands corpus conversationnels, on remarque que de multiples verbes à l'infinitif y déclenchent une liaison, ce qui pourrait être un signe que ces variétés relèvent de la distance (Koch & Oesterreicher, 2001). Cela expliquerait aussi l'émergence des nombreux pataquès en [*ʁ] qu'on relève dans le FJD. Par ailleurs, quelques pataquès en [*n] semblent pouvoir s'expliquer par la voyelle finale des M1 précédant les pataquès, nasale, à l'image de la voyelle nasale finale des M1 précédant la consonne de liaison [n].

Dans cette partie, j'ai voulu examiner quelques M1 du FJD et du corpus de tops 1, afin de déterminer si ces deux corpus chantés présentaient des spécificités. Étonnamment, je n'ai pas beaucoup observé d'adjectifs prénominaux supplémentaires en contexte de liaison, malgré le caractère distant des productions, mais il faut se souvenir qu'il s'agit de petits corpus. J'ai tout de même pu à l'occasion actualiser les relevés de Barreca (2015) pour les corpus PFC et CFPP2000. En revanche, les données concernant les M1 précédant les liaisons en //ʁ// sont plus intéressantes, puisqu'on relève de très nombreux infinitifs liaisonnants. Dans la partie suivante, je procéderai à un bilan quantitatif concernant les liaisons et les pataquès dans le FJD et le corpus de tops 1, en me concentrant tout d'abord sur les liaisons et liaisons variables.

5.4 Liaison et pataquès dans deux corpus chantés

5.4.1 Liaisons et liaisons variables dans le FJD et le corpus de tops 1

Dans leur corpus d'enfantines, Nardy et al. (2014 : 251-252) avaient noté que 348 liaisons étaient réalisées dans 473 contextes de liaison[24], soit un taux de réalisation

[24]Les auteurs avaient noté 171/174 liaisons invariables et 177/299 liaisons variables réalisées (Nardy et al., 2014 : 251-252).

global de 73,57%. En ce qui concerne les liaisons variables, ces auteurs avaient relevé un taux de réalisation de 59% (177/299 ; Nardy et al., 2014 : 252). Je m'attendais donc à rencontrer un grand nombre de liaisons lors de l'étude de nouveaux corpus chantés.

Dans l'échantillon du FJD étudié, j'ai observé un taux de réalisation global des liaisons de 68% (3704/5459)[25], et un taux de réalisation des liaisons variables s'élevant à 54% (2031/3734 ; Coutanson & Badin, sous presse)[26]. On voit donc que même s'il s'agit de textes versifiés (voir 1.2.3.4.2), toutes les liaisons ne sont pas pour autant réalisées. Si ce taux global de réalisation est en accord avec les résultats du corpus d'enfantines étudié par Nardy et al., la différence est significative concernant les taux de réalisation des liaisons variables. Toutefois, dans de premiers résultats portant sur un échantillon de 141 chansons (5h30min, 26 931 mots, 2 854 contextes de liaison), j'avais relevé un taux de réalisation global des liaisons de 58% (1654/2854) et un taux de réalisation des liaisons variables de 44% (899/2048). Entre mes premiers et mes derniers résultats pour les deux taux du FJD, la différence est significative. On se méfiera donc des observations faites sur le corpus d'enfantines, puisqu'il s'agit de très petites données, tout comme on peut se demander si les données du FJD sont suffisantes, ou si ces résultats sont encore amenés à évoluer avec davantage de données. Dans le corpus de tops 1, le taux de réalisation global des liaisons est de 54,83% (3275/5973)[27], et celui des

[25] Afin d'obtenir ces résultats, j'ai dû procéder à un nettoyage de mes données. Notamment, deux chansons occasionnaient des erreurs : il s'agit de JD_a165_28 et de JD_a062_08. J'ai supprimé les données qui ne correspondaient pas à des contextes de liaison, ainsi que les doublons qui correspondaient au même vers d'une même chanson, avec une balise temporelle initiale similaire. Je n'ai également pas retenu les sandhis externes inattendus dont la consonne n'avait pas d'équivalent parmi les consonnes de liaison (*d, *j, *l, *s et *v), les données annotées comme incertaines (j*, n*, r*, s*, t*, v* et z*), ainsi que des contextes qui ne constituaient pas des pataquès (*k, *bon* [n*] *mon*, *la* [*p] *campagne*, *a* [*r] *enfermée*, *partout* [*r] *au*, *à* [*r] *murmurer*, x [*r] *adieu*, *garde* [*r] *à*, *église* [*r] *on* et par sécurité, *moutons* [*n] *avez*.

[26] Depuis les données globales nettoyées, j'ai retiré toutes les occurrences pour lesquelles le POS du M1 était DET:POS, DET:ART, PRO:DEM, PRO:PER ou PRP:det (voir Stein, 2003 pour le jeu d'étiquettes Treetagger pour le français), contextes relevés comme invariables dans la littérature. Je n'ai pas supprimé le contexte VER + enclitique car étant annotés comme des pataquès, ils n'étaient pas compris dans ces chiffres.

[27] Suite aux résultats présentés dans Coutanson & Badin, 2021, j'ai extrait à nouveau les données du corpus de tops 1, courant 2023, en explicitant davantage le nettoyage des données qui a été effectué. Notamment, deux chansons occasionnaient des erreurs : il s'agit de 2002_01-03_07-03_Qui_est_l'exemple et de 2006_20-01_26-01_Nolwenn_ohwo_. Elles avaient été écartées par sécurité dans Coutanson & Badin, 2021. Par ailleurs, j'ai supprimé les données qui ne correspondaient pas à des contextes de liaison, présentant en particulier une balise temporelle initiale similaire. Je n'ai également pas retenu les sandhis externes inattendus dont la consonne n'avait pas d'équivalent parmi les consonnes de liaison (*a, *d, *E, *e, *f, *H, *i, *j, *l, *m, *N, *o~, *s et *ɥ), les données annotées comme incertaines (j*, n*, N*, r*, t* et z*), ainsi que des contextes qui ne constituaient pas des pataquès (*k, *n, *p, *r et *t devant initiale consonantique).

liaisons variables est de 36,76% (1516/4124)[28]. Comme nous le verrons dans la partie suivante, présenter ces chiffres pour le corpus de tops 1 n'est pas vraiment adapté, puisqu'il s'agit d'un corpus microdiachronique.

Dans le corpus variationniste PFC, le taux de réalisation global est de 43,70% (18175[29] /41587[30]), et le taux de réalisation des liaisons variables s'élève à 19,04% (4296/22568 ; Barreca, 2015 : 535). Le corpus d'enfantines, le FJD et le corpus de tops 1, trois corpus chantés, sont donc plus liaisonnants qu'un corpus conversationnel comme PFC, que ce soit au niveau global ou au niveau des liaisons variables.

J'ai ici montré que les liaisons étaient davantage réalisées, que ce soit à un niveau global ou au niveau des liaisons variables uniquement, dans le corpus d'enfantines étudié par Nardy et al. (2014), dans le FJD ainsi que dans le corpus de tops 1, sans prendre en compte le caractère diachronique de ce dernier, par rapport au corpus PFC. Observons à présent comment le phénomène de liaison a évolué en microdiachronie, de 1956 à 2017, dans le corpus de tops 1.

5.4.2 La liaison en microdiachronie dans le corpus de tops 1

Pour mon corpus de tops 1, j'ai arbitrairement divisé mes données par décennie afin de réaliser une analyse microdiachronique : 1956-1966 (chansons ayant accédé à la première place de classements entre 1956 et 1966), 1967-1976, 1977-1986, 1987-1996, 1997-2006 et 2007-2017. J'ai ajouté manuellement cette annotation dans mon tableur. Après revue de la littérature (voir 1.2.3.1.1.4), je ne m'attendais pas forcément à observer une diminution des liaisons, malgré une idée reçue assez répandue à ce sujet, puisque cela n'avait pas été observé pour les corpus HPOL (personnalités politiques) ou ESLO (variationniste), lorsque des périodes temporelles étendues avaient été prises en compte.

Au fil du temps, les liaisons sont pourtant nettement moins réalisées dans le corpus de tops 1 à un niveau global, comme cela avait été présenté dans Coutanson & Badin,

[28]Depuis les données globales nettoyées, j'ai retiré toutes les occurrences pour lesquelles le POS du M1 était DET:POS, DET:ART, PRO:DEM, PRO:PER ou PRP:det (voir Stein, 2003 pour le jeu d'étiquettes Treetagger pour le français), contextes relevés comme invariables dans la littérature.

[29]Cet effectif a été obtenu sur l'interface de recherche du corpus PFC, en ne conservant que les liaisons réalisées, les liaisons non enchaînées ainsi que les données conversationnelles, à l'aide de la requête

```
@type ("g" | "l") @liaison_presence (1 | 3).
```

[30]Cet effectif total a été obtenu sur l'interface de recherche du corpus PFC, en omettant les données incertaines et les liaisons épenthétiques et en ne conservant que les données conversationnelles, à l'aide de la requête `@type ("g" | "l") @liaison_presence (1 | 0 |3).`

2021 : 141-143. En effet, on passe de 67,76% de liaisons réalisées lors de la première décennie (643/949 ; 1956-1966), à 64,03% lors de la deuxième (614/959 ; 1967-1976), puis à 59,71% lors de la troisième (492/824 ; 1977-1986), à 53,57% lors de la quatrième (375/700 ; 1987-1996), à 46,54% lors de la cinquième (639/1373 ; 1997-2006), pour atteindre 43,84% lors de la sixième et dernière décennie étudiée (512/1168 ; 2007-2017). Il y a bien baisse puisque la différence entre les taux de la première et de la dernière décennies est bien significative.

Si on s'intéresse cette fois-ci aux liaisons variables, une diminution s'observe également : on passe de 53,91% de liaisons réalisées lors de la première décennie (352/653 ; 1956-1966), à 48,47% lors de la deuxième (300/619 ; 1967-1976), puis à 44,06% lors de la troisième (252/572 ; 1977-1986), à 32,04% lors de la quatrième (149/465 ; 1987-1996), à 26,29% lors de la cinquième (255/970 ; 1997-2006), pour arriver à 24,62% lors de la sixième et dernière décennie étudiée (208/845 ; 2007-2017). À nouveau, la baisse est bien significative. Veuillez noter au passage que les chiffres présentés dans (Coutanson & Badin, 2021 : 142) différaient quelque peu, notamment pour la dernière décennie[31], mais mes conclusions restent les mêmes, puisque c'est également une tendance à la baisse qui se dessine.

Nous avions précédemment remarqué que ces taux se rapprochaient des taux de réalisation des liaisons dans des corpus variationnistes contemporains comme PFC (Coutanson & Badin, 2021 : 142), que ce soit au niveau global ou au niveau des liaisons variables (voir taux de réalisation pour PFC en 5.4). Au contraire, les données de la première décennie sont similaires aux données du FJD, de la même façon au niveau global ou au niveau des liaisons variables (voir 5.4). Cette évolution pourrait être analysée de la sorte (Coutanson & Badin, 2021 : 142-144) :

> Ici, le changement en cours dans les chansons, à première vue diachronique, semble relever d'un alignement progressif de la langue chantée sur la langue parlée, les chansons passant de la distance à la proximité. [...] La convergence progressive de la langue chantée vers la langue parlée dans les tops 1 sur la période étudiée nous semble principalement illustrer une variation diaphasique moindre au fil du temps.

Nous avons donc pu approcher la complexité du corpus microdiachronique de tops 1, afin de dépasser une analyse générale comme proposée en 5.4. Nous venons de voir

[31] Les effectifs totaux de Coutanson & Badin, 2021 : 142 étaient erronés en ce qui concerne les liaisons variables. Cela a été corrigé dans Coutanson & Badin, sous presse.

que les liaisons avaient chuté dans les chansons à succès depuis 1956, au niveau global comme en se concentrant uniquement sur les liaisons variables. Si les chansons de la première décennie semblent liaisonner de la même façon que les autres corpus chantés, les chansons de la dernière décennie se comportent plutôt comme des données de français parlé. Cependant, ce changement diachronique apparent pourrait cacher un autre type de changement, diaphasique, la langue chantée s'alignant progressivement sur la langue parlée. Je proposerai désormais une analyse quantitative des pataquès dans le FJD et le corpus de tops 1.

5.4.3 Pataquès dans le FJD et le corpus de tops 1

Lors de notre première fouille des corpus ALIPE (voir 4.1.4.1) et PFC (voir 4.1.4.2), nous avions proposé des taux de réalisation de pataquès pour ces corpus, selon les éléments retenus comme pataquès et uniquement en corrigeant les données bruitées, sans recherche de données silencieuses. Dans ALIPE, les pataquès étaient réalisés à hauteur de 0,87% (66/7598), davantage par les enfants (2,40% ; 55/2296) que par leurs parents (0,21% ; 11/5302), et en excluant les reprises enfantines de ces derniers leur taux était abaissé à 0,11% (6/5302 ; cf 4.1.4.1). Concernant PFC, les pataquès n'étaient produits qu'à hauteur de 0,11% (57/53861), davantage en lecture (0,27% ; 33/12193) qu'en conversation (0,06% ; 24/41668 ; cf 4.1.4.2). Il est intéressant de noter que le taux de réalisation des parents d'ALIPE hors reprise de pataquès enfantins et celui des locuteurs de PFC en conversation sont similaires. S'agissant des chansons de tradition orale, je m'attendais à y rencontrer davantage de pataquès.

C'est en effet le cas puisque dans le FJD, 5,67% (328/5787[32]) de pataquès se laissent apercevoir (Coutanson & Badin, sous presse). Le FJD semble donc présenter davantage de pataquès que les corpus PFC et ALIPE, et même davantage que les productions des enfants du corpus ALIPE. Plusieurs facteurs pourraient expliquer ce taux élevé : en premier lieu, certains locuteurs du FJD sont analphabètes :

> Certains chanteurs disent ne savoir ni lire ni écrire, ce qui fait que dans les années 1960, il est encore possible d'enregistrer en France des témoins anal-

[32]Concernant l'effectif total, si les pataquès en contexte de liaison ne posent pas problème, les pataquès hors contexte de liaison amènent une complexité puisqu'ils ne sont pas produits en contexte de liaison. Afin d'éprouver la rareté du phénomène, il faut se demander quel est notre point de référence. J'observe donc les pataquès réalisés au regard de l'ensemble des contextes de liaison mais aussi des contextes où des pataquès sont réalisés, comme le font des corpus comme PFC et ALIPE où lors de l'annotation de la liaison, de telles occurrences sont prises en compte. D'autres solutions existent, notamment celle de conserver comme point de référence toutes les frontières de mots devant initiale prévocalique.

> phabètes, n'ayant jamais reçu d'instruction, puisque tous les enfants n'ont pas immédiatement bénéficié d'une école de secteur suite aux lois Ferry. [...] [Les chanteurs du FJD] ne subissent pas le poids d'une norme prenant l'écrit comme référence et ces pataquès perdurent. [...] [L]a récitation au sein du groupe de pairs semble prévenir toute intervention d'auditeurs plus lettrés sur ces productions (Coutanson & Badin, sous presse).

Par ailleurs, le FJD est un fonds très liaisonnant, présentant bon nombre de liaisons variables (voir 5.4.1), ce qui pourrait encourager l'émergence de nombreux pataquès. Enfin, dans l'échantillon de 141 chansons du FJD mentionné en 5.4.1, j'avais préalablement noté un taux de réalisation de 3,55% (105/2959). À nouveau, la différence avec le nouveau taux observé est significative, ce qui peut nous questionner sur la taille du corpus nécessaire pour obtenir un taux représentatif des pataquès en son sein.
Par contre, ces observations ne peuvent pas être réitérées pour un autre corpus chanté, le corpus de tops 1, puisqu'on y observe seulement deux pataquès (2/5975 ; 0,03% ; cf exemples 300 et 301), ce qui n'est pas vraiment étonnant pour un corpus de l'industrie du disque, notamment pour des titres ayant été classés numéro 1. On s'attend en effet à ce que ceux-ci soit dûment vérifiés avant leur sortie, étant donné que le pataquès reste un élément linguistique dévalorisé socialement et qu'il existe un enjeu pour le chanteur en termes d'image, et donc commercial.

(300) *SGR : pour qu'il sourie [*t] en s'endormant.
(Tops 1, Charles Aznavour, *La Mamma*, 1964 [1963])

(301) *SGR : lui dit viens donc [*t] à Paris ma jolie !
(Tops 1, Georgette Plana, *Riquita*, 1968 [1968])

Ainsi, on constate davantage de pataquès dans les chants traditionnel (FJD) que dans un corpus conversationnel (PFC), dans des données de lecture (PFC), dans des données acquisitionnelles (ALIPE) ou encore dans un corpus de chansons de l'industrie du disque à succès (corpus de tops 1). Ceci n'est pas propre aux corpus chantés, comme l'indiquent les résultats pour le corpus de tops 1. Dans le point suivant, nous observerons de plus près les divers ordonnancements des consonnes de liaison et de pataquès dans le FJD.

5.4.4 Ordonnancement des consonnes de liaison et de pataquès dans le FJD

Nous avions au début de ce manuscrit de thèse mentionné les divers ordonnancements des consonnes de liaison qui avaient été relevés dans la littérature (voir 1.2.2.1.3.3). Par manque de temps, je n'analyserai ici que les données du FJD, mais il serait intéressant de se pencher également sur les données du corpus de tops 1.

Pour commencer, pour le premier ordonnancement, soit la part de réalisation potentielle de chaque consonne de liaison dans l'ensemble des contextes de liaison, l'ordonnancement /z/ > /t/ > /n/ avait été observé pour les corpus PFC (Mallet, 2008 : 213) et NCCFr (Adda-Decker et al., 2012 : 122). Dans le FJD, on trouve /z/ (45,56% ; 2487/5459) > /t/ (28,27% ; 1543/5459) > /n/ (21,45% ; 1171/5459) > //ʁ// (4,21% ; 230/5459) > /p/ (0,33% ; 18/5459) = /k/ (0,18% ; 10/5459). Les taux sont très proches de ce qui a été noté pour PFC et le NCCFr (voir figure 1.5). Ainsi, si dans les observations suivantes nous observons des différences, cela ne peut être dû qu'aux réalisations, puisque la part potentielle de chaque consonne de liaison ne fluctue pas.

Concernant le second type d'ordonnancement, soit la part de réalisation de chaque consonne de liaison dans l'ensemble des liaisons réalisées, l'échantillon du FJD retenu s'organise de la sorte : /z/ (43,66% ; 1617/3704) > /t/ (28,37% ; 1051/3704) = /n/ (26,46% ; 980/3704) > //ʁ/// (1,46% ; 54/3704) > /p/ (0,05%, 2/3704). /k/ pour sa part est absent du corpus, malgré les nombreux archaïsmes qu'on y trouve. Il est assez surprenant que les taux des réalisations des consonnes /t/ et /n/ soient similaires. Cela pourrait être dû à une présence plus importante de liaisons en /t/ dans ce corpus, qui doivent être en grand nombre parmi les liaisons variables.

Le dernier type d'ordonnancement possible concerne la part de réalisation de chaque consonne de liaison dans l'ensemble des réalisations potentielles de cette même liaison. Dans le corpus de Malécot (1975 : 168) ainsi que dans les corpus PFC (Mallet, 2008 : 213) et NCCFr (Adda-Decker et al., 2012 : 122), l'ordonnancement était systématique pour les trois consonnes les plus fréquentes : /n/ > /z/ > /t/. Pour le FJD, je relève l'ordonnancement suivant : /n/ (83,69% ; 980/1171) > /t/ (68,11% ; 1051/1543) = /z/ (65,02% ; 1617/2487) > //ʁ// (23,48% ; 54/230) = /p/ (11,11% ; 2/18). À nouveau, /z/ et /t/ ont un comportement similaire. Il ne se contentent pas d'être aussi prégnants dans l'espace auditif, mais ils sont aussi liaisonné de façon aussi importante dans leurs contextes de liaison respectifs. Par ailleurs, les liaisons en //ʁ// sont davantage réalisées dans le FJD que dans le corpus PFC (1,27% ; 12/942 ; Mallet, 2008 : 213). Ceci pourrait être un indicateur du caractère distant (Koch & Oesterreicher, 2001) d'un cor-

pus.

Enfin, concernant les consonnes de pataquès, j'avais observé dans mon mémoire de master 2 l'ordonnancement suivant : /z/ (54,73% ; 110/201) > /t/ (33,33% ; 67/201) > //ʁ// (9,95% ; 20/201) > /n/ (1,49% ; 3/201) = /k/ (0,5% ; 1/201) (Coutanson, 2017 : 73). Pour l'échantillon du FJD étudié, l'ordonnancement est le suivant : /t/ (46,95% ; 154/328) = /z/ (40,85% ; 134/328) > //ʁ// (9,45% ; 31/328) > /n/ (2,74% ; 9/328). Les pataquès en [*p] et [*k] sont ici absents des chansons étudiées. Entre ces deux observations, qui ne se fondent pas systématiquement sur les mêmes données, la part de pataquès en [*z] est comparable, mais dans l'échantillon du FJD la part de pataquès réalisés en [*t] est supérieure à celle observée dans mes données de master.

Ici, je souhaitais voir si le comportement des liaisons dans le FJD confirmait nos connaissances en matière d'ordonnancements des contextes de liaison, ou si le fonds présentait des spécificités à ce sujet. Pour commencer, le premier type d'ordonnancement a été confirmé, ce qui semble montrer que les corpus ont une organisation similaire en matière de liaison. En ce qui concerne le deuxième type d'ordonnancement, le FJD se comporte un peu différemment, puisque les liaisons en /t/ et /n/ y sont réalisées à part égale. Ensuite, concernant le troisième type d'ordonnancement, le FJD diffère à nouveau des autres corpus, puisque /t/ y est davantage réalisé, mais aussi //ʁ//. Enfin, concernant l'ordonnancement des consonnes de pataquès, [*t] est davantage réalisé que dans mes observations de master.

Ce bilan quantitatif des phénomènes de liaison et de pataquès s'avère ainsi conséquent : j'ai décrit dans quelle mesure le FJD et le corpus de tops 1 présentaient des liaisons et des liaisons variables, mais aussi des pataquès, tout en comparant ces résultats aux taux observés dans le corpus PFC. Nous avons également pu observer l'évolution des liaisons et liaisons variables dans le corpus de tops 1 de 1956 à 2017. Enfin, les divers ordonnancements des consonnes de liaison du FJD ont pu être comparés à ceux d'autres corpus. Dans la partie suivante, nous aurons l'occasion de mettre à profit la transcription complète des chansons du FJD, afin de voir si certaines réalisations de liaisons permettent d'expliquer l'apparition de quelques pataquès.

5.5 Lien entre les phénomènes de pataquès et de liaison dans le FJD et le corpus de tops 1

5.5.1 Origine des pataquès dans la construction NOM singulier + *et* pour le FJD

Dans le FJD, on observe huit pataquès en [*z] dans la construction ADJ/NOM singulier + *et*, comme on peut le voir dans les exemples 302 et 309.

(302) *SGR : où allez vous jolie [*z] et bell(e) que vous marchez si lentement ?[33] (JD_a060_29, 43810, Virginie Granouillet)

On peut émettre l'hypothèse que c'est la construction ADJ/NOM pluriel + *et* qui influence ces réalisations, puisqu'à un niveau plus abstrait, ces deux constructions sont comprises dans la construction ADJ/NOM + *et*. Pour vérifier cette hypothèse, il faudrait que la construction plurielle soit liaisonnante dans le corpus, et on s'attendrait à ce qu'elle le soit dans une proportion suffisante pour avoir une influence analogique.

J'ai donc conservé toutes les occurrences apparaissant entre un M1 et la conjonction *et* du FJD, puis supprimé toutes celles pour lesquelles le M1 était un adverbe, un déterminant, un numéral, un pronom, une préposition ou un verbe, ne conservant que des adjectifs, des entités nommées et des noms, après lesquels une marque de pluriel [z] peut apparaître en liaison. J'ai également supprimé les données annotées comme incertaines, et je n'ai pas conservé les interjections propres au chant commençant par *et* annotées sous forme de blocs (*et* + *rou* + *piou* + *piou* + *piou* + *piou* + *piou* + *piou@si*).

Les liaisons sont réalisées à hauteur de 32,4% (44/136) au sein de la construction ADJ/Nom propre (NAM)/NOM pluriel + *et* dans le FJD, ce qui semble assez important pour influencer la construction au singulier. Je n'ai pas accès aux occurrences au singulier hors contexte de liaison, puisqu'elles n'ont pas été annotées. Je ne peux donc pas indiquer le taux de réalisation des pataquès dans ce contexte. Toutefois, il faut souligner qu'au singulier, [*z] est en concurrence avec les consonnes de liaison [t] (dans *nuit et jour* voire *la nuit et le jour*, cf 5.2.3), [n] (dans *le matin et le soir*, cf 4.1.3.3) et [r] (dans *le monsieur et la dame*, cf 4.1.3.3), sans oublier que /z/ n'a pas toujours une valeur plurielle (on observe *doux et malin* ainsi que *flux et reflux* au singulier, cf 5.2.3). Les liaisons dans la construction ADJ/NAM/NOM + *et*, en contexte de liaison et de pataquès, sont réalisées à hauteur de 25,2% (54/214) en [(*)z], de 5,1% en [t] (11/214),

[33] Le *vous* n'a pas ici de valeur plurielle.

de 0,9% en [n] (2/214) et de 0,5% en [r] (1/214). C'est donc la consonne de liaison la plus fréquente dans ce contexte, /z/, qui est utilisée en cas de pataquès.

Je vais terminer en mettant en avant les exemples au pluriel (exemples 303, 304, 305, 306[34], 307 et 308) et au singulier avec pataquès (exemples 309 et 310) d'un même lemme, *père*, apparaissant dans ce contexte, afin d'illustrer le phénomène étudié. On note trois non-réalisations (exemples 303, 304 et 305) et six réalisations (306, 307 et 308) de liaisons en contexte pluriel, pour trois pataquès en [*z] en contexte singulier (exemples 309 et 310). Des influences à différents niveaux pourraient coexister : en plus d'une influence de la construction ADJ/NAM/NOM pluriel + *et* sur la construction ADJ/NAM/NOM singulier + *et*, la construction *pères* + *et* pourrait influer sur la construction *père* + *et*.

(303) *SGR : fillett(e)s ne fait(e)s pas comme moi écoutez vos pèr(e)s [oz] et mères !
(JD_a043_09, 43810, Marie Soulier)

(304) *SGR : fill(e)s ne fait(e)s pas comme moi craignez vos pèr(e)s [oz] et vos mères !
(JD_a075_01, 43810, Virginie Granouillet)

(305) *SGR : fill(e)s faites pas comme moi craignez vos pèr(e)s [oz] et vos mères !
(JD_a165_13, 43800, Aimé Bouyolle)

(306) *SGR : craignez vos pères [z] et mères !
(JD_a354_03, 43800, Marie Archer)

(307) *SGR : pères [z] et mèr(e)s qui avez enfants.
[...]
*SGR : ils font pleurer pères [z] et mères.
[...]
*SGR : pères [z] et mèr(e)s qui avez enfants.
[...]
*SGR : ils font pleurer pères [z] et mères.
(JD_a347_12, 43200, Séraphine Chaize)

(308) *SGR : les pères [z] et mèr(e)s xxx vos enfants.
(JD_a350_03, 43200, Séraphine Chaize)

(309) *SGR : mon père [*z] et ma mère i(l)s sont allés aux champs.
[...]

[34] Les exemples 303, 304, 305 et 306 correspondent au même vers issu des versions de chanteurs différents.

*SGR : mon père [*z] et ma mère i(l)s sont allés aux champs.
(JD_a165_27, 43800, Germaine Varenne)

(310) *SGR : bonjour père [*z] et mère frèr(e) et soeur et parents.
(JD_a359_05, 43800, Mme Rostin)

Je me suis ici interrogée sur l'origine des pataquès en [*z] dans le contexte ADJ/NAM/NOM singulier + *et*. /z/ étant la consonne la plus fréquente dans le contexte ADJ/NAM/NOM + *et*, il semble logique que ce soit cette consonne qui apparaisse en cas de pataquès. L'origine des pataquès pourrait être analogique, provenant notamment du même contexte au pluriel, où les liaisons sont réalisées de façon conséquente. Je me suis enfin tournée vers un contexte lexical spécifique, après le lexème *père*, et il est possible que les influences entre constructions se produisent à des niveaux de granularité différents. Nous allons à présent nous pencher sur un autre cas d'étude, en nous demandant quelle pourrait être l'origine des pataquès dans la construction X + *ensemble*.

5.5.2 Origine des pataquès dans la construction X + *ensemble* pour le FJD

Dans le FJD, on note cinq pataquès en [*z][35] et un pataquès en [*r] devant le mot *ensemble*, comme on peut le voir dans les exemples 311 et 312.

(311) *SGR : ta mèr(e) n'est pas contente quand ell(e) nous voit [*z] ensemble.
(JD_a287_05, 63550, Marie Fontenille)

(312) *SGR : le chaudronnier s'en sont couchés [*r] ensemble mais d'amour ils parlai(e)nt.
(JD_a155_02, 43810, Maria Drevet)

On peut cette fois-ci émettre l'hypothèse qu'ailleurs dans le corpus, des M1 liaisonnent en [z] ou [r] avant le M2 *ensemble* et que les chanteurs ont retenu cette information sous forme de chunks consonne de liaison + *ensemble* (z + *ensemble*, r + *ensemble*), chunks n'étant pas toujours remobilisés de façon adéquate. Afin de vérifier cette hypothèse, il faudrait pouvoir observer ailleurs dans le corpus les consonnes de liaison [z] et [r] avant le mot *ensemble*, et on s'attendrait à ce que la première se note de façon plus importante étant donné que les pataquès en [*z] apparaissent dans davantage de contextes lexicaux et syntaxiques.

[35] Quatre sont issus de la même chanson, JD_a287_05, et parmi ces occurrences, on trouve deux doublons.

J'ai donc conservé toutes les occurrences apparaissant entre un M1 et le M2 *ensemble*. Les liaisons et pataquès sont réalisés à hauteur de 65,9% (27/41) en /z/, de 4,9% (3/41) en /t/ et de 4,9% (3/41) en //ʁ// dans ce contexte. Ainsi, on observe bien des liaisons en [z] avant le M2 *ensemble* (22/41), d'une façon qui semble assez conséquente pour devenir la consonne de pataquès préférentielle dans ce contexte. Les liaisons en [z] apparaissent après des verbes conjugués à la 1.PL (21/22 liaisons réalisées, 27/31 contextes de liaison, cf exemple 313), et dans un cas après le pronom *nous* (exemple 314). Il n'est donc pas surprenant que les cinq pataquès en [*z] apparaissent après des verbes conjugués. On observe également une liaison en [r] (exemple 315) et deux liaisons en [t][36] (cf exemple 316). On pourrait alors se demander ce qui a déclenché le choix entre [*r] et [*t] pour le pataquès de l'exemple 312. C'est assez simple : celui-ci se situe après un participe passé en -[e], et nous avions vu en 1.3.5.1.3 que c'était un contexte favorable à l'émergence de liaisons en [r], d'autant plus que devant le M2 *ensemble*, la seule liaison en [r] (exemple 315) correspond bien à une liaison après un verbe à l'infinitif.

(313) *SGR : nous causerons [z] ensemble.
(JD_a159_10, 43800, André-Joseph Aguilhon)

(314) *SGR : marions nous [z] ensemble c'est un(e) a(ssurance).
(JD_a043_21, 43810, Marie Soulier)

(315) *SGR : son père le maria croyant d(e) rester [r] ensemble.
(JD_a156_06, 43800, Rosine Veyrac)

(316) *SGR : et la première nuit qu'ils en furent [t] ensembl(e) et la première nuit.
(JD_a061_11, 43810, Virginie Granouillet)

Nous venons donc de nous interroger sur l'origine d'un deuxième type de pataquès dans le FJD, les pataquès produits devant le M2 *ensemble*. Nous avons vu que dans ce contexte, on observe de nombreuses liaisons en [z] dans le fonds, liaisons qui apparaissent principalement après des verbes conjugués à la première personne du pluriel. De ce fait, les pataquès produits dans la majorité des contextes lexicaux et syntaxiques sont en [*z] et émergent eux aussi après des verbes conjugués. Dans ce contexte, des liaisons en [t] et [r] sont aussi produites. Le dernier pataquès, en [*r], s'explique par sa présence après un participe passé. Nous nous pencherons dorénavant sur un dernier cas d'étude, concernant les contextes postverbaux.

[36] Les deux réalisations sont un doublon issu d'une même chanson.

5.5.3 Origine des pataquès dans la construction VER conjugué + X pour le FJD et le corpus de tops 1

En contexte postverbal (VER conjugué + X), deux consonnes de liaison sont en concurrence : /z/ et /t/. Dans les exemples ci-après issu de Coutanson & Badin, sous presse, on remarque que /z/ peut émerger après des premières et deuxièmes personnes (exemple 317) et /t/ après des troisièmes personnes (exemple 318), comme l'avaient aussi relevé Morin et Kaye (1982 : 325) ou encore Tranel (1987 : 169). J'avais préalablement proposé de considérer la première comme une marque de participant au discours et la seconde comme une marque de non-participant au discours (Coutanson, 2017 : 80). En terme de glose, la première pourrait être annotée comme une marque de *Speech Act Participant* (SAP). Dans le cadre d'une linguistique fondée sur l'usage, on s'attendrait à ce que les auditeurs extraient des régularités de leur input, et réinvestissent ces connaissances en production. Il serait donc logique de trouver préférentiellement, lors de la réalisation de pataquès dans ces contextes, des consonnes de pataquès [*z] et [*t] associées aux bonnes personnes, comme des pataquès en [*z] après des premières et deuxièmes personnes, à l'image de l'exemple 319.

(317) (JD_a168_18, 63550, Marie Fontenille)

nous = couch-er-ons	*ensemble*
nu = kuʃ-ər-ɔ̃	z-ɑ̃sɑ̃blə
SBJ.1.PL = coucher-FUT.IND-1.PL	1-ensemble

'Nous coucherons ensemble […]'

(318) (JD_a063_30, 43810, Virginie Granouillet)

le = plus = jeune	*des = trente*	*chantait*	*une = chanson*
l-ə = ply = ʒœnə	de = tʁɑ̃tə	ʃɑ̃t-ɛ	t-ynə = ʃɑ̃sɔ̃
DEF-M = SUP = jeune	ASS : DEF.PL = trente	chanter-IMPF.IND.SG	3-INDEF.F = chanson

'Le plus jeune des trente chantait une chanson'

(319) (Coutanson, 2016 : 98)

je = reviendrai	*au = pays*
ʒə = ʁəvjɛ̃-dʁ-ɛ	z-o = pɛi
SBJ.1.SG = revenir-FUT.IND-1.SG	1-à : DEF.M = pays

'Je reviendrai au pays'

J'ai donc repris mes données nettoyées (voir note de bas de page 25 p.263) issues du FJD, et je n'ai conservé que les occurrences pour lesquelles le POS du M1 est un

verbe conjugué[37]. J'ai par la suite effectué une annotation manuelle, sur mon tableur, de la personne à laquelle est conjugué chaque verbe en M1 : 12 pour les premières et deuxièmes personnes, et 3 pour les troisièmes personnes. Les résultats de cette étude sont présentés dans le tableau 5.1.

Mon hypothèse de départ est infirmée, puisqu'on note des pataquès en [*t] après des premières et deuxièmes personnes, à hauteur de 3,07% (13/423), et des pataquès en [*z] après des troisièmes personnes (comme dans l'exemple 45 p.102), à hauteur de 4,14% (36/870). On pourrait alors se demander si une tendance se laisse observer, où les pataquès en [*z] seraient majoritaires après des premières et deuxièmes personnes, et les pataquès en [*t] majoritaires après des troisièmes personnes. Dans le premier cas, ce n'est pas attesté, puisque [*z] (1,42% ; 6/423) et [*t] (3,07% ; 13/423) sont produits dans une même proportion après ces personnes. Ces contextes étant inférieurs numéralement (le FJD présente 423 contextes de liaison après des premières et deuxièmes personnes contre 870 contextes de liaison après des troisièmes personnes), les données restent peut-être trop petites pour que la différence soit significative. Dans le second cas cependant, l'hypothèse pourrait être confirmée, puisque davantage de [*t] (10,46% ; 91/870) que de [*z] (4,14% ; 36/870) sont réalisés dans ce contexte. Mais il faut rester très prudents, puisque les liaisons en [*t] en contexte postverbal, après des troisièmes personnes (70,80% ; 616/870), sont réalisées de façon beaucoup plus importante que les liaisons en [*z] dans ce contexte, après des premières et deuxièmes personnes (46,57% ; 197/423), ce qui pourrait encourager la réalisation de pataquès en [*t]. Par ailleurs, rappelons-nous que tous les /t/ de troisième personne sont étymologiques, contrairement à certains /z/ après des premières ou deuxièmes personnes (cf 1.2.3.1.2.2), ce qui pourrait amener les chanteurs à produire des pataquès en [*t] après des troisièmes personnes. Enfin, avec un corpus de chants de tradition orale, des [*t] annotés comme des pataquès pourraient être encore des archaïsmes préservés en contexte de liaison.

Pour résumer, en contexte postverbal, la liaison /z/ apparaît après des premières et deuxièmes personnes, et la liaison /t/ après des troisièmes personnes. La morphémisation des consonnes de liaisons n'ayant pas abouti, chaque consonne ne s'observe

[37] Je n'ai donc pas tenu compte des données annotées comme des participes présents (VER:ppre), des participes passés (VER:pper) et des infinitifs (VER:infi) par Treetagger (cf jeu d'étiquettes de Treetagger pour le français ; Stein, 2003). J'ai constaté que certains verbes conjugués avaient très rarement été annotés comme des noms. Je n'ai pas procédé à une réannotation manuelle et ces données n'ont pas été prises en compte lors de l'analyse.

Personne	**1S, 2S, 1P, 2P**		**3S, 3P**	
(Non)-réalisation	**%**	**n**	**%**	**n**
x	46,57	197/423	70,80	616/870
ox	48,94	207/423	14,60	127/870
***z**	1,42	6/423	4,14	36/870
***t**	3,07	13/423	10,46	91/870

TABLEAU 5.1 : Réalisation de liaisons et de pataquès postverbaux dans le fonds Jean Dumas (Coutanson & Badin, sous presse)

pas forcément après les personnes adéquates en cas de pataquès. Aucune tendance générale ne ressort n'ont plus des données, à l'exception des pataquès en [*t] qui sont produits préférentiellement après des troisièmes personnes, plutôt que des pataquès en [*z]. Pour autant, difficile de savoir si les chanteurs ont perçu une subtilité morphologique, si cela résulte de dynamiques fréquentielles, d'une influence de l'étymologie, ou encore d'archaïsmes préservés en liaison.

Dans cette partie, nous avons tenté d'expliquer pourquoi trois types de pataquès émergeaient : les pataquès dans la construction NOM singulier + *et*, les pataquès dans la construction X + *ensemble* et enfin les pataquès dans la construction VER conjugué + X. Parfois, les données de liaison du corpus ont permis d'expliquer l'apparition de certains pataquès, allant dans le sens d'une appréhension des unités linguistiques sous forme de constructions. Certains cas restent obscurs, mais cela ne nous amène pas forcément à rejeter l'approche constructionniste, puisqu'il faut se rappeler que des constructions à différents niveaux sont en concurrence. Par exemple, si un pataquès en [*z] apparaît dans *voit* [*z] *ensemble* comme dans l'exemple 311, après un verbe conjugué à la troisième personne où sont produites normalement des liaisons en /t/, il y a en réalité concurrence entre les constructions VER.3 + X et X + *ensemble*.

Dans ce cinquième et dernier chapitre, nous avons pu fouiller les deux corpus principaux sur lesquels se fonde cette thèse, soit le FJD et le corpus de tops 1. Les trois premières parties concernaient généralement des analyses qualitatives, alors que les deux dernières étaient davantage quantitatives. Au niveau qualitatif, j'ai essayé de me concentrer sur quelques spécificités du FJD et du corpus de tops 1. Le premier est le plus riche, puisqu'il s'agit d'un corpus de chansons de tradition orale, pouvant présenter certains archaïsmes, absents de corpus contemporains. Au niveau quantitatif, les deux corpus se sont révélés très intéressants pour renouveler les connaissances linguistiques

en matière de liaison et de pataquès.

Conclusion

En guise de conclusion générale, je reviendrai sur l'ensemble des éléments qui ont été abordés dans le cadre de cette thèse de doctorat, en essayant de mettre en avant les apports de ce travail de recherche.

Dans un premier chapitre consacré à un état de l'art portant sur la liaison et le pataquès, j'ai tout d'abord rappelé en 1.1 qu'il s'agissait de phénomènes de sandhi externes consonantiques, mais qu'une définition implicite du mot comme mot graphique pouvait avoir une incidence sur les contextes considérés comme des sites potentiels de liaison, et que ce choix n'était pas toujours explicitement mentionné lors de l'étude du phénomène. En 1.2.1, nous avons ensuite vu que les définitions proposées du phénomène de liaison n'étaient pas toujours neutres, puisqu'elles pouvaient laisser transparaître un cadre théorique. Cela a également été l'occasion de rappeler que les liaisons n'apparaissent pas uniquement devant initiale vocalique mais aussi devant certaines glissantes, et j'en ai profité pour recenser les différents termes qu'on rencontre dans la littérature pour désigner le mot précédant et le mot suivant la liaison, ainsi que celui qui se réfère au contexte dans lequel une liaison peut émerger. Enfin, j'estime qu'il faut réserver le terme de *liaison* à l'étude du français ou d'autres langues romanes, sans l'étendre à davantage de phénomènes morphophonologiques. En 1.2.2.1.1, j'ai remobilisé les pertinents arguments avancés par Morin (2005) allant à l'encontre d'un rôle anti-hiatus pour les liaisons, arguments qui comme nous l'avons vu proviennent de la poésie et ont été repris dans les travaux (post-)génératives. J'ai indiqué que l'émergence de liaisons non étymologiques aurait pu constituer un argument de poids allant dans le sens d'une fonction anti-hiatique, mais que cet argument a vite été remis en cause par le manque de systématicité du phénomène. En 1.2.2.1.2, j'ai rappelé que les liaisons enchaînées avaient surtout été étudiées chez les personnalités politiques, mais qu'elles restaient erratiques dans les grands corpus de parole quotidienne. Par la suite, en 1.2.2.1.3.1, j'ai retenu /p, t, n, z, ʁ/ comme consonnes de liaison en français standard

pour les corpus contemporains, /k/ correspondant à un sandhi interne (ou à une liaison prononcée [g] dans de l'écrit oralisé) et /l/ et /v/ à deux sandhis externes, le premier n'étant pas assez systématique pour être considéré comme une liaison et le second ne faisant pas alterner la présence et l'absence d'une consonne. En 1.2.2.1.3.2, nous avons constaté qu'en français, à l'exception de quelques harmonies consonantiques ou de différentes prononciations du phonème /ʁ/, on observait peu de variation allophonique lors de la réalisation des consonnes de liaison, contrairement au francoprovençal par exemple, langue romane dans laquelle ce phénomène se rencontre aussi. Comme mentionné en 1.2.2.1.3.3, on peut être perplexe face aux multiples ordonnancements des consonnes de liaison rencontrés dans la littérature. Pourtant, l'objectif de cette partie était de montrer qu'il pourrait y avoir plus de consensus qu'il n'y paraît : tout d'abord, comme l'avait ébauché Mallet (2008), il existe trois types d'ordonnancements possibles, présentant trois réalités à ne pas confondre. Ensuite, bien que l'ordonnancement le plus commun semble varier selon les corpus, il se pourrait que cela soit dû à une présence plus ou moins importante de liaisons variables en leur sein. En 1.2.2.1.4, j'ai énuméré les différents cas où une liaison pouvait avoir une influence sur le segment la précédant. Dans le cas le plus extrême, pour quelques mots monosyllabiques, seule la consonne de liaison subsiste. Quant au contexte droit, les liaisons n'apparaissent pas uniquement devant des mots à ‹h›- disjonctifs, mais aussi dans d'autres contextes que j'ai soulignés. Concernant le statut morphologique des consonnes de liaison, deux analyses sont possibles (morphème ou consonne du radical ; voir 1.2.2.2.1). J'ai proposé en 1.2.2.2.2 un résumé des différentes possibilités de recoupement entre allomorphe libre masculin, allomorphe lié masculin et allomorphe féminin (flexion), tout en rappelant un recoupement possible avec le radical utilisé en dérivation ainsi que le caractère défectif de la majorité des adjectifs masculins singuliers en liaison. En 1.2.2.2.3, j'ai cité les différents rôles fonctionnels attribués aux consonnes de liaison dans la littérature. J'ai argumenté qu'une morphémisation de ces consonnes pourrait être révélée par la présence de pataquès ou par une réalisation différentielle des liaisons selon des critères morphologiques, mais qu'elle ne pouvait être complète puisque le phénomène se produisait uniquement en liaison et que la marque restait redondante. J'ai enfin mentionné que certains auteurs traitaient les liaisons avant des enclitiques comme des cas d'allomorphisation. En 1.2.2.2.4, j'ai recensé les diverses neutralisations (en genre et en nombre) qui peuvent survenir au niveau du M1 en liaison, la neutralisation s'étendant parfois à tout le syntagme. C'est la forme longue des M1 qui est adoptée en liaison, forme qui peut être semblable à l'oral à la forme féminine, à la forme masculine ou à la forme plurielle. Dans le cas des déterminants unipossessifs, j'ai rappelé que la neutralisation constituait

une innovation.
Au niveau syntaxique, il ne faut pas se limiter à la catégorie morphosyntaxique des mots 1 et 2, qui est connue pour avoir une influence sur la réalisation de la liaison, mais quelquefois scruter des unités syntaxiques élargies (voir 1.2.2.3.1). En ce qui concerne les types de contextes de liaison, j'ai plaidé pour un abandon des termes prescriptifs de *liaisons obligatoires*, *liaisons facultatives* et *liaisons interdites*, notamment employés par Delattre (1947), et pour une adoption des termes descriptifs de *liaisons attestées* d'une part, subdivisées en *liaisons invariables*, *liaisons variables* et *liaisons erratiques* (termes d'Encrevé ; 1983), et de liaisons *non attestées* d'autre part. Cela a été l'occasion pour moi de montrer que les catégories de « liaisons interdites » et de liaisons erratiques ne se superposaient pas. Par ailleurs, une typologie par corpus me semble préférable, tout comme une visualisation sous la forme d'un continuum plutôt qu'en catégories distinctes pourrait l'être. En 1.2.2.3.2, je suis revenue sur les modifications principales de la typologie initiale de Delattre (1947) : plusieurs contextes classés comme « obligatoires » passent à la catégorie variables et quelques contextes considérés comme « facultatifs » ne sont pas attestés en corpus, changements qui peuvent être dus à une évolution diachronique ou aux caractéristiques du français que prenait Delattre pour référence. Je me suis en outre demandé si un contexte lexical spécifique pouvait être catégorisé différemment.
Au niveau lexical, j'ai montré en 1.2.2.4 que les consonnes de liaisons les moins fréquentes, /p/, //ʁ// et /n/, étaient les plus liées à des lexèmes particuliers. Je suis aussi revenue sur quelques relexicalisations de consonnes de liaisons en consonnes fixes initiales.
Enfin, j'ai proposé en 1.2.2.5 un petit bilan, rarement fait, compilant les connaissances actuelles concernant le phénomène de liaison sous un angle sémantique. Dans de très rares cas, une liaison peut permettre de distinguer sémantiquement deux énoncés. Au contraire, elle peut ponctuellement rendre deux énoncés indissociables, créant une ambiguïté, allant parfois jusqu'au calembour. Une relexicalisation d'une consonne de liaison peut en outre entraîner un changement sémantique.

Après avoir adopté un angle d'approche linguistique du phénomène de liaison, j'ai poursuivi mon état de l'art sous un angle sociolinguistique, avec une division selon les différentes dimensions de variation, sans occulter les interférences qui peuvent se produire entre elles.
Dans un premier temps, je me suis longuement arrêtée sur la dimension diachronique, bien souvent méconnue des chercheurs travaillant sur la liaison. J'ai tout d'abord rappelé quelles méthodes utilisaient les diachroniciens pour travailler en phonétique/phonologie en l'absence de documents sonores (cf 1.2.3.1.1.1). En 1.2.3.1.1.2, j'ai rappelé

que quatre dynamiques contradictoires auraient été à l'œuvre dans l'histoire du phénomène de liaison (syllabation ouverte, marquage morphologique, enchaînement généralisé et graphique), auxquelles j'ai proposé d'ajouter une dynamique fréquentielle ainsi qu'une dynamique analogique. Ces dynamiques peuvent se combiner ou entrer en confrontation. En 1.2.3.1.1.3, j'ai rappelé qu'à la fois des consonnes finales latines mais aussi des consonnes finales secondaires (l'étant devenu plus tardivement), pouvaient être à l'origine des liaisons actuelles, avant de revenir sur l'évolution diachronique de ces différentes consonnes finales, qui ont pu se maintenir lexicalement, uniquement en liaison ou au sein d'un régime ternaire, voire disparaître. En ce qui concerne les initiales vocaliques, contexte permettant la liaison, je suis revenue sur l'origine des contextes disjonctifs ainsi que sur les phénomènes ayant pu modifier l'initiale des mots comme les prothèses vocaliques, les agglutinations ainsi que les déglutinations. En 1.2.3.1.1.4, j'ai proposé un bilan des connaissances concernant la liaison en micro-diachronie dans les corpus oraux, rendus possibles par l'apparition de l'enregistrement sonore. Malgré certaines idées reçues selon lesquelles les liaisons seraient de moins en moins réalisées, cela n'est pas confirmé à un niveau global dans les données des corpus HPOL et ESLO. En se focalisant uniquement sur les liaisons variables, une diminution a parfois été observée, peut-être due à une fenêtre temporelle plus restreinte, les liaisons fluctuant en micro-diachronie. Des évolutions peuvent aussi être notées au sein de constructions spécifiques ou encore au cours d'une vie. Par ailleurs, les différentes dimensions de la variation sont en interaction, et certains changements diachroniques apparents pourraient cacher d'autres types de changements, par exemple diaphasique. En 1.2.3.1.2.1, j'ai rappelé qu'au niveau morphologique, les radicaux courts constituent généralement une innovation. En 1.2.3.1.2.2, j'ai retracé l'origine des marques de pluriel [z] et de personne [z] et [t] qu'on trouve en liaison, et évoqué comment ces dernières s'étaient étendues à de nouvelles personnes, à de nouveaux temps et/ou modes. Cela a également été l'occasion pour moi de rappeler que les liaisons dans les inversions sujet/verbe n'étaient pas étymologiques.

Dans un deuxième temps, en ce qui concerne la variation diamésique, j'ai collecté en 1.2.3.2.1 plusieurs exemples de graphies non standard de liaisons, intentionnelles ou non, issues de productions linguistiques variées. J'ai aussi rappelé quelles étaient les graphies possibles des différentes consonnes de liaison, et retracé l'origine de quelques graphies n'étant pas transparentes. J'ai enfin présenté le cas des liaisons devant enclitiques, qui nécessitent une graphie spécifique.

Dans un troisième temps, concernant la variation diatopique, nous avons vu en 1.2.3.3 que même parmi les invariables, certaines liaisons pouvaient faire défaut selon les va-

riétés. Le taux de réalisation des liaisons variables fluctue pour sa part entre les différentes régions francophones, ce qui s'observe au sein de constructions spécifiques. L'ordonnancement tout comme la présence de toutes les consonnes de liaison peut enfin varier selon les géolectes, quand certains mots ne sont pas liaisonnants dans chaque variété.
Dans un quatrième temps, pour ce qui est de la variation diaphasique, j'ai rappelé en 1.2.3.4.1 qu'en lecture, les liaisons étaient davantage réalisées qu'en conversation, ce qui s'observe au sein de constructions spécifiques. Par la suite, j'ai pris le temps de rassembler en 1.2.3.4.2 les quelques informations qu'on trouve dans les ouvrages de référence à propos du lien entre liaison et versification. J'ai soutenu que les phénomènes de liaison et de pataquès n'étaient pas liés à des questions de métrique. Dans la seule étude existante portant sur la liaison dans la langue chantée (cf 1.2.3.4.3), Nardy et al. (2014) ont observé des taux élevés de réalisation et peu de variation entre les différentes versions d'une même chanson, tout en se questionnant sur l'apprentissage des liaisons au sein de chansons et sur la transmission possible de celles-ci de la langue chantée à la langue parlée.
Dans un cinquième temps, concernant pour finir la variation diastratique, Hornsby (2012) a noté davantage de liaisons en lecture chez les lycéens en filière générale, issus de milieux plus favorisés, que chez les lycéens en filière professionnelle, mais aucune différence en conversation (cf 1.2.4.1). En 1.2.4.2, aucune différence n'a été observée à première vue en conversation selon le niveau d'études des locuteurs dans le corpus PFC.

Dans la suite de cet état de l'art, je me suis tournée vers le phénomène de pataquès, en réfléchissant tout d'abord en 1.3.1 à une appellation à donner au phénomène. Les termes prescriptifs ont été vigoureusement rejetés, et celui de *pataquès* a été retenu en tant qu'hyperonyme. Par la suite, j'ai longuement commenté ma quête d'une définition adéquate pour le phénomène, et j'ai tout d'abord considéré qu'il s'agissait d'un type de liaison. J'ai progressivement rejeté les définitions du phénomène comme des liaisons non normatives, en revenant au passage sur les différentes normes qui existent (cf 1.3.2.1), comme des liaisons non standard (cf 1.3.2.2), comme des liaisons non étymologiques ou non historiques (cf 1.3.2.3), puisque se posait la question du chronolecte de référence et que certaines liaisons invariables ne sont pas elles-mêmes étymologiques, comme les liaisons après *ils*, pour en venir à les définir comme des liaisons inattendues, par rapport à un critère graphique (cf 1.3.2.4), qui reste le critère le plus automatisable par l'outil informatique, même si cela nécessite d'y inclure les liaisons en [t] qui apparaissent dans les inversions sujet-verbe. En 1.3.3, j'ai passé en revue les autres pro-

ductions erratiques en matière de liaison n'étant pas des pataquès : les invariables non réalisées, les variables fréquentes non réalisées, les erratiques, les remplacements d'une consonne initiale fixe par une consonne de liaison en contexte de liaison et les suppressions d'une consonne initiale fixe prise pour une consonne de liaison hors contexte de liaison. En 1.3.4, je suis revenue sur les différents types de pataquès : les pataquès en contexte de liaison vs hors contexte de liaison, puis les pataquès paradigmatiques vs syntagmatiques.

Dans une partie consacrée au phénomène de pataquès d'un point de vue linguistique, j'ai commencé par évoquer les différentes consonnes de pataquès relevées dans la littérature : /l, n, ʁ, t, z/, voire /k, p/ (cf 1.3.5.1.1). Si j'avais observé l'ordonnancement /z/ > /t/ > //ʁ// > /n/ = /k/ sur une collection d'exemples d'une partie du FJD, Hornsby (2012 : 76) avait pour sa part noté l'ordonnancement /t/ > /z/ > /n/ lors d'un test de lecture proposé à des lycéens. Les pataquès apparaissent préférentiellement dans le contexte V#_#V (cf 1.3.5.1.2). En 1.3.5.1.3, j'ai mentionné le cas des pataquès en //ʁ// après des participes passés, où //ʁ// n'est pas une marque morphologique d'infinitif mais est déclenché par un contexte phonétique similaire.
En 1.3.5.2, j'ai notamment plaidé, par économie, pour ne pas considérer /z/ comme une marque d'impératif mais comme une marque de participant au discours.
En 1.3.5.3, j'ai longuement recensé tous les contextes morphosyntaxiques de la littérature dans lesquels des pataquès apparaissent, en illustrant le tout d'exemples. Cela a été l'occasion de revenir sur les différentes possibilités d'analyse pour les sandhis inattendus après le complémenteur *qui* (ex : *les enfants qui* [*z] *arrivent*), que je considère comme des pataquès.
En 1.3.5.4, j'ai montré que la réalisation d'un pataquès pouvait modifier le sens d'un énoncé, en mettant en évidence un exemple de calembour.

J'ai terminé cet état de l'art en abordant le phénomène de pataquès sous un angle sociolinguistique, dans un premier temps d'un point de vue diachronique (cf 1.3.6.1). Dans mes recherches de master, je m'étais questionnée sur l'origine des pataquès dans les chansons traditionnelles (FJD), en émettant l'hypothèse que certains pourraient provenir d'un chunk issu d'un autre vers de la même chanson ou d'une autre version de la chanson. Une recherche des chunks à l'échelle du corpus n'avait pas pu être menée, étant donné que celui-ci n'avait pas été transcrit.
Dans un deuxième temps, concernant la variation diamésique (voir 1.3.6.2.1), j'ai relevé et agencé de multiples graphies non standard de pataquès, nombre d'entre elles provenant de la littérature scientifique. Le phénomène n'étant pas considéré comme une production linguistique standard, il n'est habituellement pas orthographié.

Dans un troisième temps, pour ce qui est de la variation diatopique, j'ai inventorié les différents pataquès cités dans la littérature selon les variétés (cf 1.3.6.3). Si certains semblent au premier abord spécifiques à un géolecte, on se rend souvent compte qu'ils existent ailleurs mais que cela n'est jamais mentionné. Je suis allée jusqu'à rapporter certaines lexicalisations du français louisianais, qui ne sont pas complètes puisqu'on note de la variation entre les locuteurs, contrairement aux lexicalisations qu'on observe dans les créoles à base française. Pour finir, un corpus présentant de mêmes chants traditionnels peut permettre d'étudier la répartition d'un pataquès sur une aire géographique donnée.
Dans un quatrième temps, pour finir sur la variation diagénique, dans un test de lecture proposé par Hornsby (2020) à des lycéens, davantage de filles que de garçons ont réalisé des pataquès (cf 1.3.6.4).

Dans un second chapitre méthodologique, j'ai présenté à la fois la méthodologie générale mise en place dans le cadre de cette thèse de doctorat, méthodologie comprenant quatre étapes, mais aussi comment celle-ci avait été appliquée, plus spécifiquement, lors de la constitution des deux corpus sur lesquels je me suis appuyée principalement.

La première étape de cette méthodologie consistait à effectuer un va-et-vient entre l'état de l'art et le recueil d'une collection d'exemples (cf 2.1.1), terme que j'ai défini pour l'occasion. L'état de l'art permet de prendre le temps de recenser toutes les appellations données au phénomène étudié et de le définir, ce qui nécessite de s'intéresser aux phénomènes connexes afin de savoir où le borner. Le manque d'écrits scientifiques se consacrant intégralement au pataquès et les appellations multiples données à ce phénomène ont complexifié cette étape. Concernant la collection d'exemples, j'ai privilégié des exemples oraux et contextualisables. J'ai par ailleurs distingué les exemples relevés à la volée des exemples attestés, ayant recours aux deux, mais uniquement aux seconds à titre de preuve. Les exemples peuvent enfin provenir de sources diverses ou d'un seul corpus, permettant alors d'aller plus en profondeur. Par la suite, j'ai montré comment une collection d'exemples pouvait être limitative par rapport à une étude sur corpus, avant de me demander si ce dernier type d'étude était envisageable pour un phénomène rare. Pour finir, j'ai mentionné quels linguistes travaillent traditionnellement sur des données authentiques et ont besoin minimalement de collections d'exemples et maximalement de corpus, quels ont été les premiers types de corpus constitués, avant de rappeler comment l'émergence de la linguistique de corpus a été favorisée par certains progrès techniques, permettant de dépasser les collections d'exemples.
La seconde étape de cette méthodologie impliquait de choisir puis carotter des corpus

disponibles variés (cf 2.1.2). J'ai commencé par recenser de multiples paramètres à prendre en compte lors du choix d'un corpus : langue(s), corpus synchronique vs diachronique, corpus clos vs ouvert, corpus écrit vs oral (audio, vidéo ou multimodal), corpus de référence vs spécialisé ou encore corpus brut vs annoté. Cela a été l'occasion pour moi de présenter d'autres langues dans lesquelles le phénomène de liaison peut se produire. Dans le cadre de cette thèse, mes langues d'étude sont le français (principalement) mais aussi le francoprovençal. J'ai eu recours à des corpus synchroniques et diachroniques, mais uniquement oraux, souvent spécialisés dans l'étude de la liaison et enfin annotés ou bruts. Le chant, et dans une moindre mesure la parole, sont les deux modalités qui suscitent mon intérêt, j'avais donc besoin au minimum de corpus audio. Je suis finalement revenue sur les notions de bruit et de silence dans le cadre de l'étude du pataquès et de la liaison. J'ai par la suite mentionné quelques bases de données à destination des linguistes où sont référencés au niveau national (ORTOLANG, COCOON), européen (CLARIN) ou international (OLAC) des corpus déjà constitués. Dans certains cas, il ne faut pas bouder des corpus issus d'autres disciplines ou d'archives patrimoniales (portail Gallica de la BnF, INA...). De nombreux corpus restent très confidentiels, n'ayant pas été diffusés à grande échelle par les scientifiques les ayant produits. On peut enfin constituer soi-même un corpus. J'ai enchaîné sur la notion de carottage, que j'ai opposée à la notion d'échantillonnage, puisqu'il s'agit d'un échantillonnage non représentatif, visant à donner des renseignements au linguiste sur le corpus qu'il entreprend d'étudier.

La troisième étape de cette méthodologie appelait, si nécessaire, à constituer de nouveaux corpus (cf 2.1.3). Je suis revenue sur les données sur-représentées et sous-représentées dans les corpus en langue française selon Habert (2000), toutefois ces informations commencent à dater. Une première façon de constituer un nouveau corpus, la moins impliquante, est de compiler des documents existants en un corpus exclusif. Cela concernera la majeure partie des corpus écrits. J'ai exemplifié de nombreux types de documents écrits pouvant alimenter ces corpus, mais aussi quelques documents oraux. Néanmoins, l'oral invite souvent au recueil de nouveaux corpus. J'ai là aussi énuméré de nombreux types d'enregistrements qui pouvaient être captés. Je me suis par la suite concentrée sur la chaîne de traitement mise en place pour le FJD et le corpus de tops 1. Dans un premier temps, il me fallait récupérer les fichiers son et établir un plan de nommage (cf 2.1.3.1). J'ai indiqué que pour ces deux corpus je ne devais produire que des données secondaires, et que si je devais mettre en place un plan de nommage pour le corpus de tops 1, je pouvais réutiliser celui déjà instauré pour le FJD. Dans un deuxième temps, je devais préparer un fichier .cha (cf 2.1.3.2). J'ai commencé par présenter le lo-

giciel CLAN, puis je suis revenue sur le choix de son utilisation, résultant d'une décision commune à un groupe de chercheurs, avant de préciser comment je m'étais formée à ce logiciel. J'ai par la suite rapidement expliqué comment s'organisait un fichier .cha, en indiquant les informations communes aux deux corpus retenues au niveau des métadonnées. Dans un troisième temps, un alignement du son avec le texte a été effectué (cf 2.1.3.3). Dans ce point, j'ai explicité mon choix de subdiviser les chansons en différentes tires. Dans un quatrième temps vient la phase de transcription et d'annotation du document sonore. J'ai tout d'abord proposé au lecteur de se familiariser avec les règles de transcription et d'annotation propres au logiciel CLAN (cf 2.1.3.4.1). J'ai mentionné quelques annotations sur lesquelles je suis revenue (mots entièrement omis, élisions systématiques de voyelles devant voyelles, interjections propres au chant). J'ai également indiqué ne pas avoir transcrit les autres langues que le français dans les chansons présentant du code-switching, et ne pas avoir annoté les pauses et les chevauchements. Je suis ensuite revenue sur le système de codage de la liaison et du pataquès adopté par un groupe de chercheurs en acquisition (cf 2.1.3.4.2), auquel je me suis conformée, même si je l'ai adapté à mes besoins : la consonne de liaison potentielle est indiquée en cas de non-liaison et je dispose d'annotations spécifiques, une pour les pataquès et une pour les cas incertains ou les consonnes finales réalisées habituellement non prononcées. J'ai conscience que la double fonction de cette dernière annotation reste problématique, même si j'ai précisé ne pas l'utiliser lors de mes analyses. Seuls les sandhis externes inattendus hors contexte de liaison doivent être annotés au fil de la transcription. Après cela, j'ai présenté un bilan comparatif détaillé des éléments retenus par les systèmes de codage de la liaison utilisés pour les corpus ALIPE, HPOL1, PFC, ainsi que pour le corpus de tops 1 et le FJD de cette thèse. J'ai en outre présenté un outil informatique d'aide à l'annotation pour ces deux derniers corpus, créé par Flora Badin, qui permet d'obtenir des fichiers ELAN enrichis (alignement automatique au token, étiquetage morphosyntaxique, lemmatisation), logiciel que j'ai présenté pour l'occasion. J'ai évoqué deux problèmes rencontrés lors de l'utilisation de l'outil : des décalages dans l'alignement son/texte ainsi que la suppression ponctuelle de certains tokens, engendrant du silence lors du repérage des contextes de liaison. J'ai mentionné d'autres enrichissements qui auraient pu être apportés, certains étant automatisables. J'ai ensuite expliqué comment l'outil repérait les contextes de liaison, fonctionnalité qui empêche les oublis, en soumettant au lecteur la liste de derniers caractères de M1, la liste de premiers caractères de M2, ainsi qu'une liste d'exclusion pour les M1 du corpus de tops 1 qui ont été fournies à l'outil. Adda-Decker et al. (2012) ont également voulu automatiser ce repérage, malgré certains choix restrictifs. Pour ma part,

j'ai une conception très large de ce qu'est un contexte de liaison, ne conservant que des critères grapho-phonétiques, par mesure de précaution lors de l'étude de nouveaux corpus (corpus non contemporains, acquisitionnels, d'apprenants du français, variétés non hexagonales...) et afin de permettre une bonne comparabilité entre les divers corpus. J'ai recommandé a minima de bien déclarer ce qui est considéré comme des contextes de liaison pour chaque étude, mais aussi de rendre publics nos corpus respectifs. L'outil de Flora Badin pré-annote également automatiquement les liaisons, ce qui permet un gain de temps, comme non réalisées par défaut pour le corpus de tops 1, et comme réalisées par défaut pour le FJD. Lors de la phase de correction manuelle, les contextes repérés à tort comme des sites potentiels de liaison sont supprimés, et l'annotation des liaisons est passée en revue. J'ai fait savoir que les liaisons apparaissant lors de passages chantés à l'unisson posaient problème, tout en révélant un défaut de ces deux corpus, pour lesquels je reste la seule annotatrice. Je suis enfin revenue sur les essais d'automatisation de l'annotation de la liaison par Adda-Decker et al. (2012), puis Hutin et al. (2022), avec un alignement au phonème, le tout couplé d'un dictionnaire de prononciation indiquant de potentielles liaisons, voire de modèles de phones pour repérer les consonnes en question sur un spectrogramme. Avec la deuxième étude, il y a eu une progression vers le repérage automatique de pataquès, à l'exception de [*r], mais le taux d'erreurs reste supérieur aux annotations correctes. Dans un cinquième temps, je réécoutais et vérifiais chaque fichier .cha produit (cf 2.1.3.5), en deux temps.
La quatrième étape de cette méthodologie consistait à analyser les corpus les plus pertinents (cf 2.1.4), révélés lors de la phase de carottage et de première fouille des données (deuxième étape), ou fraîchement constitués (troisième étape). Dans le cadre de cette thèse, je me suis penchée plus en détail sur le corpus de tops 1 et le FJD, tout juste transcrits et annotés, que ce soit pour effectuer une analyse qualitative en tirant profit des corpus comme réservoirs d'exemples, ou pour effectuer une analyse quantitative.

En ce qui concerne plus spécifiquement le traitement du corpus de tops 1 (cf 2.2.1), j'ai tout d'abord indiqué comment avaient été sélectionnés les titres : ils devaient présenter un passage en français et avoir été classés numéro un entre 1956 et 2017. J'ai fait part de réserves à l'égard des titres retenus, puisque des discordances s'observent selon les sites ayant archivé le classement du Top Singles pour la France. Je suis revenue sur l'histoire de ce classement et sur le type de titres qui y étaient comptabilisés, ainsi que sur le classement antérieur à 1984 proposé par le site Infodisc. On retiendra qu'une grande prudence est nécessaire vis-à-vis des titres composant le corpus de tops 1, qui pourrait être amélioré une fois les classements stabilisés, bien qu'il reste représentatif d'une époque. Les audios pour leur part ont été obtenus depuis Youtube. J'ai ensuite

présenté le plan de nommage que j'ai mis en place pour le corpus, en mentionnant un regret, celui d'avoir opté pour un format en année_jour-mois plutôt que année_mois-jour, avant d'illustrer le tout par quelques exemples. Par la suite, j'ai préparé des fichiers .cha à associer à ces sons (cf 2.2.2). J'ai recensé les métadonnées complémentaires pour ce corpus (nom du chanteur et lieu de naissance), et comment j'avais accédé à ces informations, en faisant mention d'un risque d'erreur. Il fallait pour continuer obtenir les paroles de chaque chanson et aligner le son avec le texte (cf 2.2.3). Pour gagner du temps, les paroles proviennent de sites collaboratifs où elles sont déposées par des internautes. Elles ont été recherchées manuellement, sans être aspirées, et j'ai utilisé un découpage traditionnel en vers. J'ai poursuivi en transcrivant et en annotant le corpus (cf 2.2.4). Les paroles ont été contrôlées et adaptées au format de transcription et d'annotation de CLAN. Je suis revenue sur l'annotation du schwa et sur celle de la liaison, immédiate pour ce premier corpus. Une phase de réécoute et de vérification clôture cette chaîne de traitement (cf 2.2.5).

Concernant le traitement cette fois-ci du FJD, j'ai pour commencé énuméré les différents documents qu'on trouve pour chaque chanson sur la BIPO (cf 2.3). S'agissant du téléchargement du fichier son et de la mise en place d'un plan de nommage (cf 2.3.1), chaque audio provient de la BIPO et j'ai réutilisé le plan de nommage établi lors de de la mise à disposition du corpus sur le site. L'irrégularité de ce dernier s'est révélée très incommode, et il a bien souvent fallu ruser pour s'approcher au mieux d'un plan de nommage homogène. J'ai donné quelques exemples à titre d'illustration. L'absence d'un fichier descriptif recensant l'ensemble des chansons du corpus et leur hiérarchisation hétérogène ne m'ont pas facilité la tâche. Par la suite, des fichiers .cha ont été associés à chaque audio (cf 2.3.2). J'ai opté pour l'utilisation d'un fichier modèle au fil du temps, comprenant des métadonnées pré-renseignées, et présenté les informations additionnelles retenues pour ce corpus (nom, métier et lieu de vie du chanteur). J'ai également décrit les trois rôles que pouvaient jouer les participants à la performance. L'étape suivante consistait en un alignement son/texte (cf 2.3.3), ici fondé sur les pauses saillantes à l'écoute et les tours de parole, ce qui était plus simple mais qui pourrait compromettre certaines réutilisations du corpus. J'ai évoqué deux exceptions et comment celles-ci avaient été résolues. Pour continuer, j'ai transcrit et annoté les données (cf 2.3.4). La transcription concerne uniquement les passages chantés par les chanteurs. Je suis revenue sur l'annotation du schwa, puis très longuement sur celle de la (non-)réalisation de certaines consonnes du FJD. Cela a été l'occasion de présenter quelques spécificités du fonds, comme des non-réalisations de consonnes internes aujourd'hui prononcées sous l'influence de graphies étymologisantes, une prononciation

en [j] à la place de la séquence [lj] du français standard contemporain, pour laquelle j'ai relevé de nombreux exemples et des archaïsmes dans la prononciation de consonnes finales (soit non prononcées alors qu'elles le seraient aujourd'hui, soit prononcées alors qu'elles ne le seraient plus aujourd'hui), ce qui a été illustré par des exemples d'apparence inédits au vu de la littérature. J'ai également proposé un bilan concernant le fonctionnement des consonnes finales de mots ayant un statut particulier : *quant/quand*, *plus* et *tous* pour lesquels j'ai, le cas échéant, utilisé le postcode [x*] tout en mettant en avant quelques archaïsmes, ainsi que les nombres cardinaux, pour lesquels j'ai distingué quatre types de fonctionnement. J'ai aussi abordé le cas du traitement des doublons de ce fonds. J'ai précisé pour finir que je n'avais pas océrisé les documents existants lors de cette étape. Pour terminer, chaque transcription annotée a été vérifiée au fil de deux réécoutes (cf 2.3.5). À l'issue de la première, une comparaison a été faite avec la transcription des paroles de Jean Dumas.

Dans un troisième chapitre essentiellement consacré aux deux corpus principaux sur lesquels s'appuie cette thèse de doctorat, le corpus de tops 1 et le fonds Jean Dumas, je suis tout d'abord revenue, en guise d'introduction, sur les Archives de la Parole, les premières archives sonores en langue française à avoir été enregistrées (cf 3.1). J'ai passé en revue les différents types de captations que les instigateurs du projet avaient cru bonnes d'immortaliser, et montré en quoi ce projet était novateur.

J'ai enchaîné sur une présentation du FJD, en expliquant d'abord comment j'avais choisi ce corpus (cf 3.2.1). Pour me référer aux chansons présentes dans ce fonds, j'ai opté pour l'appellation de *chansons de tradition orale*, sans retenir celles de *chansons populaires* ou de *chansons traditionnelles*, avant de souligner les critères qui les opposent aux chansons écrites. J'ai indiqué avoir commencé à collecter des pataquès écrits, avant de me tourner vers l'oral, afin de garantir leur authenticité et d'envisager chaque texte au sein d'une performance. Dans cette quête, j'ai aussi privilégié les chansons de collectage aux reprises par des chanteurs contemporains. Cela a été l'occasion pour moi de présenter la Base inter-régionale du patrimoine oral (BIPO), où est mis à disposition du public le FJD. J'ai aussi expliqué pourquoi je n'avais pas recueilli un nouveau corpus ethnomusicologique, avant d'engager une discussion concernant les réutilisations commerciales problématiques de certains enregistrements ethnomusicologiques.
J'ai poursuivi en présentant la personne de Jean Dumas, folkloriste à l'origine du recueil du fonds (cf 3.2.2). Je suis ensuite revenue sur l'organisation de ce fonds, dont le dépôt sur la BIPO n'est pas achevé. Celui-ci est subdivisé en cinq sous-fonds. L'un d'eux, lui-même divisé en cinq corpus que j'ai décrits, comprend les collectes sonores

de Jean Dumas, réalisées entre 1958 et 1962 dans le centre de la France. L'échantillon que j'étudie dans cette thèse comprend 498 chansons (18h23min, 89853 mots, 5787 contextes de liaison possibles).

J'ai pour finir commenté la démarche de Jean Dumas, en commençant par les points forts de celle-ci (cf 3.2.3.1). Parfois à contre-courant de ses contemporains, ce collecteur n'hésitait pas à faire du terrain et avait saisi de bonne heure tout l'intérêt d'enregistrer les performances de chanteurs traditionnels, mais aussi l'urgence qu'il y avait à le faire. La pratique des chants de tradition orale était en effet amenée à être redéfinie, dans une période de grands changements entre scolarisation universelle, exode rural et mutations technologiques. Ces dernières ont aussi facilité le travail de collecte de Jean Dumas, qui disposait d'une automobile et avait acquis un magnétophone. Par ailleurs, son poste de professeur lui permettait de bénéficier de vacances scolaires, qui s'ajoutaient aux jours fériés obtenus grâce aux avancées sociales, et il pouvait compter sur le soutien de son épouse, qui l'assistait dans ses collectes. Jean Dumas refusait les intermédiaires et a utilisé la méthode boule de neige pour rencontrer des interprètes, tout en tâchant de rester au plus près de leurs conditions habituelles de chant, même si cela n'a pas toujours été possible. J'ai ensuite avancé quelques exemples pour montrer que les données secondaires de Jean Dumas restaient insuffisantes, que ce soit sur le plan linguistique ou ethnomusicologique, d'où l'intérêt de l'accès aux données primaires. J'ai informé le lecteur des critiques faites à l'encontre de la notation musicale occidentale dans le champ de l'ethnomusicologie, tout en rappelant que l'enregistrement ne permet pas de rendre une performance dans son entièreté. J'ai critiqué au passage la recherche de linéarité de notre collecteur dans les chansons qu'il enregistrait, pour atteindre des versions idéalisées qui vont pourtant à l'encontre des pratiques des chanteurs traditionnels. Pour continuer, j'ai souligné la richesse des fiches rédigées par Jean Dumas pour chaque interprétation, où il apporte des informations sociologiques concernant les chanteurs et laisse transparaître une bonne conscience des diverses variations phonétiques à l'œuvre, ce que j'ai illustré à l'aide de nombreux exemples. J'ai terminé en présentant cette fois-ci les faiblesses de la démarche de ce collecteur (cf 3.2.3.2). Notamment, on peut être surpris par la rareté des chansons en langues régionales dans le fonds. Jean Dumas économisait par ailleurs à la fois les bobines, coûteuses, ne cherchant pas à préserver la langue parlée des chanteurs en marge de la performance, mais aussi la surface écrite. Nous avons également vu à l'aide de quelques exemples que son éthique des relations interpersonnelles laissait parfois à désirer. De plus, j'ai rappelé le caractère solitaire de sa démarche, ayant peu connaissance d'entreprises similaires en son temps et étant resté incompris des institutions. J'ai enfin relevé quelques défauts

propres aux enregistrements effectués, comme leur son généralement faible ou la présence d'éléments perturbateurs dus aux conditions écologiques de recueil.

Avant de passer à la présentation du corpus de tops 1, j'ai souhaité faire une transition dont l'objectif était de montrer en quoi les deux objets chanson que sont les chants de tradition orale et ceux issus de l'industrie du disque se distinguent. Notamment, ces derniers ont un auteur connu, leur durée est réduite, la co-présence spatio-temporelle du chanteur et de son auditoire et leur co-construction de l'œuvre n'ont habituellement plus cours, la consommation devient plus individuelle et massive et enfin davantage de liberté est accordée à l'auditeur.

J'ai donc clos ce chapitre par une présentation du corpus de tops 1 (cf 3.3). J'ai tout d'abord expliqué pourquoi avoir constitué ce corpus : il m'a permis de m'entraîner à la transcription et à l'annotation avec CLAN, tout en présentant un intérêt scientifique, à la fois pour accéder à de nouvelles données micro-diachroniques concernant la liaison dans la parole publique, mais aussi pour proposer aux linguistes un corpus inédit en langue chantée.
Je suis ensuite revenue sur les caractéristiques techniques de ce corpus (21h 27min, 368 chansons, 107 550 mots, 5914 contextes de liaison possibles). J'ai indiqué ne pas avoir procédé à un échantillonnage afin d'obtenir un corpus représentatif, puisque le corpus était complet, comprenant tous les titres recherchés. Pour continuer, j'ai passé en revue quelques caractéristiques sociologiques concernant les chanteurs de ce corpus : il s'agit en majorité d'hommes, ayant entre 20 et 35 ans. J'ai par ailleurs cité les dix chanteurs les plus représentés dans la tête du classement. J'ai aussi examiné les titres retenus, dont le nombre varie chaque année, puis fait remarquer qu'un petit nombre de titres se maintenait longtemps en tête du hit parade, quand un grand nombre n'y faisait qu'un très bref passage. Cette description du type de chansons et de chanteurs ayant accédé à la première place du tops 1 pour la France, bien que je ne prenne en compte que les titres francophones et que les titres retenus appellent à la prudence, ne semble jamais avoir été proposée auparavant. Ces informations semblent pourtant d'intérêt afin de comprendre le champ musical français.

Dans les quatrième et cinquième chapitres, j'ai présenté les résultats de cette thèse, couplés à une discussion. Le quatrième chapitre concerne en premier lieu les résultats du carottage, voire d'une première fouille des corpus.

J'ai commencé, pour ce qui est du carottage (cf 4.1), par énumérer les corpus choisis tout en revenant sur la raison de leur sélection et en indiquant à quels corpus j'avais dû renoncer. J'ai poursuivi par une présentation de ces corpus (cf 4.1.1), en proposant no-

tamment un tableau récapitulatif de leurs caractéristiques, les pyramides des âges des populations concernées, puis un tableau présentant ce que je nomme leur état d'avancement, permettant de savoir si je pouvais immédiatement les étudier. Dans un premier temps, concernant le FJD (cf 4.1.1.1), j'ai rappelé que n'étant pas transcrit autrement que manuellement ni systématiquement annoté pour la liaison et le pataquès, son analyse était compromise. J'ai aussi remarqué que les chanteurs y étaient particulièrement âgés. Dans un deuxième temps, pour ce qui est du corpus PFC (cf 4.1.1.2), j'ai rappelé qu'une partie des données n'étaient pas libres d'accès puisque n'étant pas anonymisées, à moins de disposer des accès à une base de données réservée, et qu'une partie du corpus n'était ni transcrite ni annotée pour la liaison. J'ai aussi montré que malgré une pyramide des âges plus équilibrée pour ce corpus, la part des locuteurs ayant une vingtaine d'années était plus importante, ce pourquoi j'ai émis une hypothèse. Dans un troisième temps, concernant le corpus ALIPE (cf 4.1.1.3), les enfants sont issus de couples hétérosexuels de trentenaires appartenant aux classes moyenne ou moyenne-supérieure. J'ai rappelé l'âge des enfants aux différents temps de recueil. Dans un quatrième temps, l'ALAVAL (cf 4.1.1.4) est pour sa part difficilement exploitable, puisqu'un retour sur les données primaires est impossible, que la transcription est phonétique, que la liaison et le pataquès n'y sont pas annotés bien que présents dans la transcription, et enfin que les données ne peuvent être exportées autrement qu'au format PDF. La population d'étude est équilibrée hommes/femmes et selon les villages, mais dans le cadre du travail sur une langue en danger, les locuteurs y sont relativement âgés. Enfin, dans un cinquième temps, pour ce qui est du fonds Charles Joisten (cf 4.1.1.5), le substrat des chanteurs traditionnels est principalement le francoprovençal, mais le lieu de collecte reste trop proche de celui du FJD. Par rapport à cet autre fonds, davantage d'informations sociologiques concernant les chanteurs faisaient défaut, et l'état d'avancement des deux fonds était similaire. Je n'ai donc pas étudié plus en détail la composition de la population d'étude de ce fonds.

J'ai ensuite présenté la méthodologie de carottage, voire de première fouille des données mise en place (cf 4.1.2), en commençant par cette dernière. Celle-ci a pu être envisagée pour les corpus PFC et ALIPE, déjà annotés pour la liaison. J'ai tout d'abord rappelé que PFC pouvait être fouillé de deux façons : grâce à une interface de recherche en ligne, que j'ai ici utilisée, ou via le programme Dolmen, tandis qu'ALIPE pouvait l'être à l'aide de CLAN. S'agissant du carottage, je souhaitais prélever des carottes de vingt minutes dans chaque corpus, uniformiser les transcriptions et annotations (exception faite de l'ALAVAL), avant d'y observer les phénomènes de pataquès, de liaison, voire d'autres sandhis externes. J'ai poursuivi en envisageant quelques critères de sélection

pour les carottes de chaque corpus.

Pour continuer, j'ai présenté les résultats de la phase de carottage (cf 4.1.3), en mettant en avant notamment un tableau récapitulatif du nombre de pataquès observés lors du carottage de chaque corpus.

Dans un premier temps, s'agissant de l'ALAVAL (cf 4.1.3.1), le format de l'atlas et des contraintes de temps m'ont amenée à réduire la durée de carottage à dix minutes. Je me suis questionnée quant à une possible influence de la langue de travail, le français, sur le francoprovençal lors de l'enquête, mais aussi en diachronie, puisque le français est une langue dominante. J'ai relevé quatre pataquès, en présentant les énoncés retenus lors de cette phase de carottage. Cela a été l'occasion pour moi de revenir sur le contexte *de* + NOM.PL, où un pataquès peut apparaître dans certains parlers pour le pluriel indéfini/partitif devant un nom à initiale vocalique, bien que les formes préconsonantiques ne comprennent que la préposition *de* et pas de déterminant. J'ai aussi mentionné trois autres sandhis externes non étymologiques qu'on observe dans l'ALAVAL : [k], [g] et [p]. J'ai argumenté les raisons pour lesquelles je ne les considérais pas comme des pataquès. J'ai par ailleurs relevé un contexte liaisonnant en francoprovençal qui ne l'est pas en français (devant le M2 *tante*), avant de m'arrêter sur le comportement des numéraux, tout particulièrement sur le cas de *neuf*. Bien qu'une analyse systématique de la liaison en francoprovençal aurait beaucoup à nous apporter, j'ai terminé en appelant à la prudence, afin de ne envisager le phénomène dans cette langue romane au regard de notre connaissance du phénomène en français.

Dans un deuxième temps, concernant le corpus ALIPE (cf 4.1.3.2), j'ai finalement conservé les productions des parents adressées à l'enfant, et indiqué de quels enregistrements provenaient les carottes, où je n'ai pas observé de pataquès. Heureusement, une première fouille des données étant possible, je pouvais vérifier si cela se confirmait à l'échelle du corpus.

Dans un troisième temps, s'agissant du FJD (cf 4.1.3.3), j'ai tout d'abord présenté les chansons retenues lors de cette étape de carottage, en montrant en quoi leur transcription et leur annotation m'avaient permis d'envisager le traitement à venir des données secondaires de ce corpus. J'ai relevé cinq pataquès au cours du carottage, ainsi qu'une liaison erratique après un nom singulier.

Dans un quatrième temps, au sujet du fonds Charles Joisten (cf 4.1.3.4), j'ai énuméré les diverses chansons ayant été prélevées, où j'ai noté deux pataquès (l'un d'entre eux reste incertain), en remarquant au passage de nombreux doublons par rapport aux chansons du FJD. Je me suis arrêtée plus longuement pour l'occasion sur le cas des [*n] qui apparaissent dans des phrases affirmatives, en proposant trois hypothèses concernant leur

origine : métanalyse d'une liaison, métanalyse de la négation ou allomorphie du pronom *en*, tout en étant peut-être favorisés par une voyelle nasale adjacente, par assimilation. Par prudence, j'ai décidé de revoir ma notation du phénomène et de ne plus utiliser l'annotation [*x], car la présence massive de ce phénomène pourrait bouleverser l'ordonnancement des consonnes de pataquès. J'ai également relevé une liaison erratique après un nom propre, et une liaison en l'absence de M1.
Dans un cinquième temps, pour ce qui est du corpus PFC, (cf 4.1.3.5), j'ai tout d'abord présenté les locuteurs dans les conversations desquels j'ai prélevé des carottes. J'ai expliqué comment j'avais converti les transcriptions au format CLAN, et quelles modifications j'avais dû apporter. J'ai constaté une absence de réalisation des liaisons après le pronom *ils* chez un locuteur âgé normand. J'ai aussi relevé quelques [*l] qui ne sont pas toujours facilement analysables, une harmonie consonantique se produisant en contexte de liaison et deux absences de liaison remarquables.
J'ai terminé en me demandant si cette étape de carottage en valait la peine. Dans le cadre de cette thèse, cette phase m'a permis de mieux savoir où borner le phénomène de pataquès, d'envisager la phase de transcription et d'annotation à venir du corpus de tops 1 et du FJD ainsi que d'observer des pataquès dans tous les corpus, en tenant compte également des résultats de la première fouille des données lorsqu'elle était possible. Toutefois, un carottage reste chronophage et son caractère non représentatif n'en font pas une étape obligatoire.

J'ai poursuivi en revenant sur les résultats de l'étape de première fouille des corpus, en commençant par le corpus ALIPE (cf 4.1.4.1). En ce qui concerne le caractère écologique des données tout d'abord, j'ai fait preuve de prudence au sujet des réalisations de Prune, qui comprennent une part non négligeable d'élicitations qui pourraient avoir engendré un nombre important de pataquès. Je n'ai pas considéré toutes les « liaisons hors contexte » codées par Loïc Liégeois comme des pataquès. Notamment, j'ai exclu les consonnes de liaison réalisées lorsque le M1 et le M2 n'appartiennent pas au même tour de parole, souvent produits par deux locuteurs différents, les consonnes de liaison réalisées lorsque le M1 est absent, les *n'* apparaissant dans des phrases affirmatives, les sandhis externes consonantiques qui ne viennent pas d'une consonne de liaison (comme le [*l] dans [mɔ̃lasjɛt]), les consonnes qui ressemblent à des consonnes de liaison mais qui remplacent en fait des consonnes n'étant pas encore maîtrisées par l'enfant, les liaisons réalisées en contexte disjonctif, des occurrences pour lesquelles je suis en désaccord avec l'annotation ainsi que quelques occurrences incertaines. Toutes les occurrences exclues ont été citées. Par ailleurs, les « liaisons hors contexte » ne constituaient pas les uniques pataquès du corpus. En ce qui concerne ces derniers, que j'ai également inva-

riablement listés, j'en ai notés dans ce corpus acquisitionnel aussi bien chez les parents que chez les enfants. Les parents répètent parfois les pataquès de leur enfant, mais apportent rarement une correction ou une reformulation dans le corpus ALIPE. J'ai aussi souligné que certains pataquès enfantins sont semblables à ceux que produisent les adultes, bien que d'autres semblent exclusivement enfantins, comme dans le contexte DET + NOM. Pour finir, j'ai proposé un bilan quantitatif de cette première fouille du corpus, où pour rappel je n'ai pu agir que sur les données bruitées. Les pataquès y restent rares (0,87% ; 66/7598), mais on en trouve davantage dans les productions enfantines (2,4% ; 55/2296) que dans les productions des adultes (0,2% ; 11/5302).

J'ai ensuite enchaîné avec les résultats de la première fouille du corpus PFC (cf 4.1.4.2). Si j'ai bien recensé à nouveau tous les pataquès relevés, je n'ai cette fois-ci pas pris le temps de présenter les cas écartés. Je suis revenue sur l'annotation des « liaisons épenthétiques » utilisée pour le codage du corpus, qui ne permet pas d'atteindre tous les pataquès, tout en renvoyant à des données bruitées. Une bonne connaissance des contextes d'apparition préférentiels pour les pataquès, notamment lexicaux, m'aura permis d'accéder à davantage d'occurrences. Par exemple, suite à la tâche de lecture, les réalisations de centaines de locuteurs peuvent être comparées pour de mêmes contextes de liaison, et on constate que plusieurs locuteurs produisent de mêmes pataquès dans les mêmes contextes. Certains pataquès semblent communs à une zone géographique particulière, ce que j'ai illustré dans deux cas à l'aide de cartes. Pour ce qui est des données conversationnelles, j'ai recensé tous les contextes morphosyntaxiques dans lesquels j'ai relevé des pataquès. J'ai aussi mentionné une absence surprenante d'un type de pataquès, en émettant une hypothèse concernant cette absence. Enfin, j'ai proposé un bilan quantitatif de cette première fouille du corpus, où à nouveau je n'ai pu agir que sur les données bruitées. Les pataquès y restent rares (0,11% ; 57/53876), et on en relève davantage en lecture (0,27% ; 33/12193) qu'en conversation (0,06% ; 24/41668).

Le cinquième et dernier chapitre de cette thèse doctorale se concentrait en second lieu sur quelques résultats relatifs à la fouille du corpus de tops 1 et du FJD.

Je suis tout d'abord revenue sur quelques cas de dénasalisations surprenantes dans le FJD, en m'intéressant dans un premier temps à des dénasalisations de voyelles finales de M1 hors contexte de liaison, devant un nom débutant en [n] (cf 5.1.1). J'ai indiqué que cela pourrait être un indice de métanalyse de la part des locuteurs, qui verraient cette consonne initiale du M2 comme une consonne de liaison.
Dans un second temps, je me suis penchée sur un particularisme du FJD : la fermeture et dénasalisation de la voyelle correspondant au graphème ‹un›, lorsqu'elle est en finale

d'un M1 en contexte de liaison (cf 5.1.2). Cela concerne notamment le déterminant indéfini masculin *un*, qui prend en liaison la même forme que son équivalent féminin. Cette prononciation est relativement répandue dans le FJD, puisqu'elle y est réalisée à hauteur de 58% (90/157). Cette découverte semble intéressante pour deux raisons : premièrement, nous sommes en présence de données attestant une fermeture et une dénasalisation de la voyelle /œ̃/ en /y/ en contexte de liaison, ce qui n'est généralement pas abordé dans la littérature ; deuxièmement, cela constitue un exemple supplémentaire de neutralisation au niveau du genre du M1 en liaison. J'ai indiqué que l'analyse perceptive pourrait évoluer en analyse acoustique, même s'il faudrait faire preuve de prudence puisqu'il s'agit de langue chantée.

Dans la partie suivante, je me suis intéressée à quelques productions erratiques de liaisons dans les deux corpus, en observant pour commencer quelques non-réalisations surprenantes de liaisons dans des contextes qui seraient aujourd'hui considérés comme invariables (cf 5.2.1). Cela a été l'occasion pour moi de revenir sur la notion d'erreur de performance. Dans un cadre génératiste, de telles erreurs sont vues comme agrammaticales, alors que dans une linguistique fondée sur l'usage, elles s'intègrent au système qu'est la langue, et peuvent même s'y trouver expliquées.

J'ai ensuite examiné quelques liaisons réalisées en contexte disjonctif (cf 5.2.2). Je suis tout d'abord revenue sur les cas des toponymes et du gentilé *Hollande*, *Hollandais* et *Hongrie*, qui sont liaisonnés dans le FJD. Cela a été l'occasion de se renseigner sur leurs étymologies respectives, et nous avons vu que *Hongrie* avait une origine latine qui ne devrait pas déclencher de disjonction. Certaines graphies historiques enfin semblent indiquer que la disjonction n'a pas toujours eu lieu pour l'ensemble de ces mots. Il en va de même pour le nom propre *Henriette*, qui est toujours liaisonné dans le fonds, malgré son origine germanique. Ces réalisations vont à l'encontre des recommandations de Fouché (1956 : 467-460) en matière de liaison. Tous ces exemples viennent questionner nos certitudes au sujet des M2 habituellement considérés comme disjonctifs.

J'ai enfin analysé des attestations surprenantes de liaisons (cf 5.2.3), allant à l'encontre de ce qui est habituellement recommandé ou constaté dans la littérature (Delattre, 1947 ; Fouché, 1956). Il s'agit par exemple de liaisons dans les locutions prépositives *par/pour rapport à*, après un nom ou adjectif singulier, entre le pronom *en* et la préposition *à/au*, après l'adverbe *demain*, entre un nom pluriel et les adverbes *en/y* ou encore entre un nom pluriel et un verbe. Toutes ces réalisations imprévues m'ont fait plaider pour l'adoption d'une définition large du contexte de liaison et pour une annotation systématique de tous les contextes, ce qui permettrait une réelle comparabilité inter-corpus.

Par la suite, je me suis intéressée au lexique employé dans certains contextes de liaison en M1, et en premier lieu aux adjectifs masculins prénominaux (cf 5.3.1). J'ai pu actualiser la première liste proposée par Barrecca (2015 : 191) pour le corpus PFC, mais les données de mes deux corpus ne se sont pas révélées très intéressantes, puisque quatre adjectifs supplémentaires seulement ont pu être notés en liaison.
En second lieu, je me suis penchée sur les M1 avant les consonnes de liaisons les moins fréquentes : /p/, //ʁ// et /n/. En effet, nous avions vu qu'habituellement, ces consonnes étaient très liées à quelques lexèmes particuliers. Les résultats pour /p/ ne se sont de même pas révélés intéressants. C'est surtout la consonne //ʁ// qui a présenté les résultats les plus spectaculaires : que ce soit dans le corpus de tops 1 ou dans le FJD, de très nombreux infinitifs sont liaisonnants, contrairement au corpus PFC qui n'en présentait qu'un en conversation, sans compter de nombreux pataquès dans des contextes variés du FJD. Ce pourrait être un signe que ces variétés relèvent de la distance. Enfin, pour la consonne /n/, seul le FJD présente quelques particularités, puisqu'on y observe quelques liaisons après des noms singuliers, ainsi que des pataquès après des lexèmes inédits, parfois à voyelle nasale finale.

Le point suivant avait pour objectif de présenter les résultats des analyses quantitatives principales de cette thèse, concernant les phénomènes de pataquès et de liaison dans le FJD et le corpus de tops 1. En ce qui concerne les liaisons et liaisons variables des deux corpus (cf 5.4.1), on peut noter qu'elles sont réalisées de façon conséquente, beaucoup plus que dans le corpus variationniste PFC, mais sans être systématiques, bien qu'il s'agisse de textes versifiés. Notamment, les liaisons invariables sont produites à hauteur de 68% (3704/5459) dans le FJD, et les liaisons variables à hauteur de 54% (2031/3734). Cela a été l'occasion pour moi de me demander si mes corpus étaient assez importants pour que les taux observés se stabilisent. Ensuite, en 5.4.2, je me suis concentrée sur le corpus de tops 1 afin d'observer l'évolution de ces taux en diachronie. De façon surprenante par rapport à nos connaissances actuelles en matière de liaison, les liaisons ont chuté drastiquement entre ces deux périodes, au niveau global (de 68% ; 643/949 à 44% ; 512/1168) ainsi qu'au niveau des liaisons variables (de 54% ; 352/653 à 25% ; 208/845). Les données de la première décennie sont plus proches de données chantées, quand les données de la dernière décennie se rapprochent de la langue parlée. Nous serions ainsi en présence d'un changement diaphasique davantage que diachronique. Par la suite, je me suis penchée plus particulièrement sur le comportement des pataquès (cf 5.4.3). J'ai relevé au passage que les données des parents du corpus ALIPE convergeraient avec les données conversationnelles de PFC, même s'il reste nécessaire d'être prudent. Le FJD présente un nombre conséquent de pataquès :

5,67% (328/5787), ce qui peut nous amener à nous demander s'il s'agit encore d'un phénomène rare au sein de ce fonds. Ce fort taux pourrait entre autres être dû à un poids moindre de l'écrit sur ces chanteurs, qui interprètent des chansons transmises oralement et qui n'ont pas tous été alphabétisés, et à la forte présence de liaisons réalisées au sein du corpus. Le corpus de tops 1 pour sa part en présente très peu (0,03%; 2/5975), ce qui n'est pas étonnant puisqu'un retour sur les données produites est possible avant la diffusion, que la réalisation de pataquès reste dévalorisée socialement et que des enjeux d'image et commerciaux entrent en jeu. Enfin, mon intérêt s'est porté sur les divers ordonnancements des consonnes de liaison et de pataquès dans le FJD (cf 5.4.4). J'ai observé, concernant le premier ordonnancement, que la part de réalisation potentielle de chaque consonne de liaison dans l'ensemble des contextes de liaison est similaire dans ce corpus à celle de PFC et du NCCFr, ce qui signifie que chaque consonne de liaison a autant de chance de pouvoir être produite dans tous ces corpus. Pour ce qui est du deuxième ordonnancement, ou part de réalisation de chaque consonne de liaison dans l'ensemble des liaisons réalisées, les liaisons en /t/ semblent plus nombreuses dans le fonds, ce qui fait qu'elles le sont autant que les consonnes en /n/. La consonne /k/ est absente du fonds, même si celui-ci présente de nombreux archaïsmes. Enfin, s'agissant du troisième type d'ordonnancement, ou part de réalisation de chaque consonne de liaison dans l'ensemble des réalisations potentielles de cette même liaison, /n/ > /t/ = /z/ peut être noté, contrairement à l'habituel /n/ > /z/ > /t/. En outre, la part des liaisons en //ʁ// (23%; 54/230) dans le FJD, très importante, pourrait être un marqueur de distance. Enfin, l'ordonnancement des consonnes de pataquès diffère des résultats obtenus au cours de mes recherches de master : j'observe /t/ = /z/ > //ʁ// > /n/ et non plus /z/ > /t/ > //ʁ// > /n/ = /k/, les pataquès en [*t] étant numéralement équivalents aux pataquès en [*z].

Pour finir, j'ai tenté de saisir le lien qui existe entre les phénomènes de pataquès et de liaison dans le FJD, pour commencer dans la construction NOM singulier + *et* (cf 5.5.1). J'ai remarqué que les liaisons étaient réalisées à hauteur de 25% (54/214) en [(*)z] dans le contexte ADJ/NAM/NOM + *et*, ce qui en fait une consonne préférentielle en cas de pataquès. Ce taux est en grande partie dû aux contextes pluriels, puisque la construction ADJ/NAM/NOM pluriel + *et* est liaisonnante à hauteur de 32% (44/136). L'influence des contextes pluriels sur les contextes singuliers pourrait se coupler d'influences à d'autres niveaux : *pères* + *et* pourrait par exemple aussi affecter son équivalent singulier, à l'image des pataquès qu'on y observe.

J'ai ensuite avancé un second exemple, touchant à la construction X + *ensemble* (cf 5.5.2). Dans ce contexte, les liaisons sont réalisées à hauteur de 66% (27/41) en /z/,

notamment après des verbes conjugués, de 4,9% (3/41) en /t/ et de 4,9% (3/41) en //ʁ//. On trouve alors cinq pataquès en [*z] en contexte verbal et un pataquès en [*ʁ] après un participe passé, cas déjà abordé en 1.3.5.1.3.
Pour finir, un dernier contexte a retenu mon attention : le contexte postverbal (cf 5.5.3, où deux consonnes de liaison, /z/ et /t/, sont en concurrence, la première apparaissant après des premières et deuxièmes personnes, et la seconde après des troisièmes personnes. En cas de pataquès, chaque consonne n'est pas forcément utilisée dans le contexte adéquat, ce qui aurait indiqué une morphémisation partielle de ces consonnes de liaison. Une tendance pourrait à la rigueur être notée dans le FJD pour les pataquès en [*t] après des troisièmes personnes, ce qui montrerait que les locuteurs ont été capables d'extraire cette information de leur input et de la réinvestir en production. Néanmoins, cette tendance pourrait n'être due qu'à la fréquence plus importante de liaisons en /t/ réalisées, au fait que cette liaison est étymologique ou encore à des archaïsmes préservés en contexte de liaison.

Annexes

I Références des énoncés sélectionnés lors du carottage de l'ALAVAL

TABLEAU 2 : Références des énoncés sélectionnés lors du carottage de l'ALAVAL

Contexte morphosyntaxique	**Référence locuteur**	**Durée de l'extrait**
1.1. Le nom des témoins (11100-Appelle)	ArbazF	00:03
1.2. L'âge des témoins (11200-Age)	ArbazM	00:01
1.3. Le lieu de naissance des témoins (11300-JeSuisNe)	BionazF	00:02
2.1.1. L'article défini féminin singulier : systèmes acasuel et bicasuel (21010-ArticleDefiniFSgPrecons)	BionazM	00:03
2.1.2. L'article défini masculin singulier : systèmes acasuel et bicasuel (21012-ArticleDefiniMSgPrecons)	ChalaisF	00:03
2.1.3. L'article défini singulier préconsonantique : synthèse (21013-ArticleDefiniSingulierSynthese)	ChalaisM	00:02
2.1.4. L'article défini pluriel : neutralisation et maintien de l'opposition masculin/féminin (21014-ArticleDefiniPluriel)	ChamosonF	00:02
2.1.5. Emploi et non-emploi de l'article défini avec les noms propres (21015-ArticleDefini + NomPropre)	ChamosonM	00:03
2.1.6. Forme et emploi du morphème pluriel devant substantif à initiale vocalique (21016-MorphemePlurielPrevoc	ContheyF	00:05

Suite du tableau 2 : Références des énoncés sélectionnés lors du carottage de l'ALAVAL

2.1.7. L'article défini précédé de «à» : les formes préconsonantiques du féminin singulier (21020-PrepA + ArticleDefiniFSgPrecons)	ContheyM	00:03
2.1.8. L'article défini précédé de «à» : les formes préconsonantiques du masculin singulier (21022-PrepA + ArticleDefiniMSgPrecons)	ÉvolèneF	00:03
2.1.9. L'article défini précédé de «à» : les formes préconsonantiques du pluriel (21024-PrepA + ArticleDefiniPl)	ÉvolèneM	00:03
2.1.11. L'article défini précédé de «de» : les formes préconsonantiques du pluriel (21064-PrepA + ArticleDefiniPl)	FullyF	00:09
3.1.1. Le pronom personnel tonique de la 1re personne du singulier (31010-Tonique1)	FullyM	00:04
3.1.2. Le clitique sujet préconsonantique de la 1re personne du singulier (31020-ClitiquePrecons1)	HérémenceM	00:02
3.1.2.1. Le clitique sujet préconsonantique de la 1re personne du singulier : fréquence d'emploi (31021-ClitiquePrecons1Stat)	HérémenceF	00:02
3.1.3. Le clitique sujet prévocalique de la 1re personne du singulier (31030-ClitiquePrevoc1)	IsérablesF	00:02
3.1.3.1. Le clitique sujet prévocalique de la 1re personne du singulier : fréquence d'emploi (31031-ClitiquePrevoc1Stat)	IsérablesM	00:03
3.2.1. Le pronom personnel tonique de la 1re personne du singulier (32010-Tonique2)	La Chapelle-d'AbondanceM	00:02
3.2.2. Le clitique sujet préconsonantique de la 2e personne du singulier (32020-Clitique2)	La Chapelle-d'AbondanceF	00:02
3.2.3. La syntaxe de la proposition interrogative à la 2e personne du singulier : formes et occurrences (32030-SyntaxeInterrogative2)	LensF	00:03
3.3.1. Le pronom personnel tonique féminin de la 3e personne du singulier (33010-Tonique3f)	LensM	00:01

Suite du tableau 2 : Références des énoncés sélectionnés lors du carottage de l'ALAVAL

3.3.2. Le pronom personnel tonique masculin de la 3e personne du singulier (33020-Tonique3m)	Les MarécottesF	00:04
3.3.3. Le clitique sujet préconsonantique féminin de la 3e personne du singulier (33030-ClitiqueSujetPrecons3f)	Les MarécottesM	00:02
3.3.3.1. Le clitique sujet préconsonantique féminin de la 3e personne du singulier : fréquence d'emploi (33031-ClitiquePrecons3fStat)	LiddesF	00:02
3.3.4. Le clitique sujet préconsonantique masculin de la 3e personne du singulier (33040-ClitiqueSujetPrecons3m)	LiddesM	00:02
3.3.4.1. Le clitique sujet préconsonantique masculin de la 3e personne du singulier : fréquence d'emploi (33041-ClitiquePrecons3mStat)	LourtierF	00:02
3.3.5. Le clitique sujet prévocalique féminin de la 3e personne du singulier (33050-ClitiqueSujetPrevoc3f)	LourtierM	00:03
3.3.5.1. Le clitique sujet prévocalique féminin de la 3e personne du singulier : fréquence d'emploi (33051-ClitiquePrevoc3fStat)	MiègeF	00:03
3.3.6. Le clitique sujet prévocalique masculin de la 3e personne du singulier (33060-ClitiqueSujetPrevoc3m)	MiègeM	00:04
3.3.6.1. Le clitique sujet prévocalique masculin de la 3e personne du singulier : fréquence d'emploi (33061-ClitiquePrevoc3mStat)	MontanaF	00:04
3.3.9. Le pronom réfléchi de la 3e personne (33210-PronomReflechi3)	MontanaM	00:04
3.3.15. Le pronom personnel objet indirect de la 3e personne (33220-PronomIndirect3)	NendazF	00:02
3.4.1. Le clitique sujet préconsonantique de la 1re personne du pluriel (34010-ClitiqueSujet4Precons)	NendazM	00:02
3.4.2. Le clitique sujet prévocalique de la 1re personne du pluriel (34020-ClitiqueSujet4Prevoc)	OrsièresF	00:02

Suite du tableau 2 : Références des énoncés sélectionnés lors du carottage de l'ALAVAL

3.4.3. Le pronom régime de la 1re personne du pluriel (34030-PronomRegime4)	OrsièresM	00:03
3.5.1. Le pronom personnel tonique de la 2e personne du pluriel (35010-Tonique5)	St-JeanF	00:02
3.5.2. Morphologie du clitique sujet de la 2re personne du pluriel : les formes prévocaliques (35020-ClitiqueSujetPrevoc5)	St-JeanM	00:04
3.5.3. Syntaxe du clitique sujet de la 2re personne du pluriel (35030-ClitiqueSujet5Syntaxe)	SavièseF	00:02
3.5.4. Le pronom régime de la 2e personne du pluriel (35040-PronomRegime5)	SavièseM	00:04
3.5.5. Syntaxe de la proposition interrogative à la 2e personne du pluriel : formes et occurrences (35050-SyntaxeInterrogative5)	SixtF	00:02
3.6.1. Le clitique sujet préconsonantique féminin de la 3e personne du pluriel (36020-ClitiqueSujetPrecons6f)	SixtM	00:06
3.6.2. Le clitique sujet préconsonantique masculin de la 3e personne du pluriel (36030-ClitiqueSujetPrecons6m)	TorgnonF	00:03
3.6.2.1. Le clitique sujet préconsonantique masculin de la 3e personne du pluriel : fréquence d'emploi (36031-ClitiquePrecons6mStat)	TorgnonM	00:18
3.6.3. Le clitique sujet prévocalique féminin de la 3e personne du pluriel (36040-ClitiqueSujetPrevoc6f)	TroistorrentsF	00:02
3.6.3.1. Le clitique sujet prévocalique féminin de la 3e personne du pluriel : fréquence d'emploi (36041-ClitiquePrevoc6fStat)	Troistor-rentsM	00:04
3.6.4. Le clitique sujet prévocalique masculin de la 3e personne du pluriel (36050-ClitiqueSujetPrevoc6m)	Val-d'IlliezF	00:03

Suite du tableau 2 : Références des énoncés sélectionnés lors du carottage de l'ALAVAL

3.6.4.1. Le clitique sujet prévocalique masculin de la 3e personne du pluriel : fréquence d'emploi (36051-ClitiquePrevoc6mfStat)	Val-d'IlliezM	00:02
3.6.6. Le pronom régime de la 3e personne du pluriel (36200-PronomRegime6)	VouvryM1	00:05
4.1.1.1. «aller» (41110-AllerInf)	VouvryM2	00:02
4.1.1.2. «vêler» (41120-VelerInf)	ArbazM	00:06
4.1.2.3. «donner» (lat. bajulare) (41230-BalyiInf)	ArbazF	00:04
4.1.3.1. «dormir» (41310-DormirInf)	BionazF	00:04
4.1.3.2. «venir», «tenir» et composés (41320-VenirInf)	ChalaisF	00:09
4.2.1.1. *asseditatu «assis», masculin singulier et pluriel (42110-AsseoirPPm)	BionazM	00:03
4.2.1.2. *asseditata «assise», féminin singulier et pluriel (42120-AsseoirPPf)	ChalaisM	00:04
4.2.1.3. *asseditatu/*asseditata «assis, assise» : l'accord du participe passé conjugué avec «être» (42130-AsseoirPPsynth)	ChamosonF	00:02
4.2.1.4. Le participe passé de «compter» (42140-CompterPP)	ChamosonM	00:03
4.2.2.2. «donné» (lat. *bajulatu) (42220-BalyiPP)	ContheyF	00:07
4.2.3.3. Le participe passé de «pouvoir» (42330-PouvoirPP)	ContheyM	00:03
4.2.3.4. Le participe passé de «voir» (42340-VoirPP)	ÉvolèneF	00:04
4.2.3.5. Le participe passé de «faire» (42350-FairePP)	ÉvolèneM	00:05
4.2.3.6. Le participe passé de «mordre» (42360-MordrePP)	FullyF	00:02
4.2.3.7. Syntaxe : L'accord au féminin du participe passé des verbes réfléchis précédés d'un complément d'objet indirect (42400-SyntaxeAccordPPReflechiCI)	HérémenceF	00:02
5.1.1.1. «j'achète» (51110-AcheterPres1)	FullyM	00:02
5.1.1.2. «je bâille» (51120-BaillerPres1)	HérémenceM	00:02

Suite du tableau 2 : Références des énoncés sélectionnés lors du carottage de l'ALAVAL

5.1.1.3. «je travaille» (51130-TravaillerPres1)	IsérablesF	00:02
5.1.1.4. «je dors» : morphologie du radical (51140-DormirPres1Rad)	IsérablesM	00:01
5.1.1.5. «je dors» : morphologie de la désinence (51141-DormirPres1Des)	LensF	00:02
5.1.1.6. «je vais» : morphologie du radical (51150-AllerPres1Rad)	La Chapelle-d'AbondanceF	00:01
5.1.1.7. «je vais» : morphologie de la désinence (51151-AllerPres1Des)	La Chapelle-d'AbondanceM	00:01
5.1.1.8. «je fais» : morphologie du radical (51160-FairePres1Rad)	LensF	00:01
5.1.1.9. «je fais» : morphologie de la désinence (51161-FairePres1Des)	LensM	00:02
5.1.1.10. «je prends» : morphologie de la radical (51165-PrendrePres1Rad)	LiddesF	00:03
5.1.1.11. «je prends» : morphologie de la désinence (51166-PrendrePres1Des)	LourtierF	00:05
5.1.1.12. «je viens, je me souviens» : morphologie du radical (51171-VenirPres1Rad)	LiddesM	00:20
5.1.1.13. «je viens, je me souviens» : morphologie de la désinence (51172-VenirPres1Des)	MiègeF	00:03
5.1.1.14. «je suis» (51175-EtrePres1)	LourtierM	00:02
5.1.1.15. «j'ai» (51180-AvoirPres1)	MiègeM	00:03
5.1.1.16. «je vois» (51190-VoirPres1)	MontanaM	00:02
5.1.2.1. «tu vois» (51220-VoirPres2)	NendazM	00:02
5.1.3.1. «il/elle/on voit» (51320-VoirPres3)	MontanaF	00:02
5.1.3.4. «il/elle fait» (51350-FairePres3)	NendazF	00:02
5.1.3.5. «Venir» et composés : 3e personne singulier de l'indicatif présent (51370-VenirPres3)	OrsièresF	00:02
5.1.4.1. «nous travaillons : la nature du radical (51420-TravaillerPres4Rad»)	OrsièresM	00:02

Suite du tableau 2 : Références des énoncés sélectionnés lors du carottage de l'ALAVAL

5.1.4.2. «nous travaillons : emplacement de l'accent tonique et nature de la désinence (51422-TravaillerPres4Des»)	St-JeanF	00:02
5.1.4.3. «nous sommes» (51440-EtrePresent4)	St-JeanM	00:04
5.1.4.4. «nous avons» (51450-AvoirPresent4)	SavièseF	00:06
5.1.4.5. «nous allons» : les formes de la base verbale (51460-AllerPresent4Rad)	SavièseM	00:01
5.1.4.6. «nous allons» : les formes de la désinence (51470-AllerPresent4Des)	TorgnonF	00:01
5.1.4.7. «nous mettons» (51480-MettrePres4)	TorgnonM	00:04
5.1.4.8. «nous voyons» (51490-VoirPres4)	Troistor-rentsM	00:02
5.1.5.1. Verbes en -ARE à radical non palatalisant : «vous achetez» (51510-AcheterPres5)	TroistorrentsF	00:01
5.1.5.2. «vous allez» (51515-AllerPres5)	Val-d'IlliezF	00:01
5.1.5.2. Verbes en -ARE à radical palatalisant : «vous travaillez» (51520-TravaillerPres5)	Val-d'IlliezM	00:03
5.1.5.3. Verbes en -ARE à radical palatalisant : «vous mangez» (51530-MangerPres5)	ArbazF	00:02
5.1.5.4. «vous êtes» (51540-EtrePres5)	VouvryM1	00:03
5.1.5.5. «vous voyez» (51550-VoirPres5)	ArbazM	00:02
5.1.5.6. «vous venez» (51560-VenirPres5)	BionazF	00:01
5.1.5.6. «asseyez-vous» : La désinence de la 2e personne du pluriel de l'impératif d'un verbe en -are à radical non palatalisant (51570-Sasseoir !5)	VouvryM2	00:01
5.1.6.1. «ils/elles achètent» (51605-AcheterPres6)	BionazM	00:05
5.1.6.3. «ils/elles mangent» (51620-MangerPres6)	ChalaisM	00:03
5.1.6.4. «ils/elles voient» (51630-VoirPres6)	ContheyM	00:03
5.1.6.5. «ils/elles viennent» (51640-VenirPres6)	ContheyF	00:04
5.1.6.6. «ils/elles sortent» : présence de formes inchoatives (51650-SortirPres6)	ChamosonM	00:05
5.1.6.7. «ils/elles vont» (51660-AllerPres6)	ChamosonF	00:03
5.1.6.8. «ils/elles font» (51670-FairePres6)	ChalaisF	00:06

Suite du tableau 2 : Références des énoncés sélectionnés lors du carottage de l'ALAVAL

5.1.6.9. «ils/elles boivent» (51672-BoirePres6)	ÉvolèneF	00:02
5.1.6.10. «ils/elles ont» (51680-AvoirPres6)	ÉvolèneM	00:01
5.1.6.11. «ils/elles sont» (51685-EtrePres6)	FullyF	00:03
5.2.1.1. Verbes en -ARE à radical non palatalisant : «je rentrais» (et substituts) (52110-RentrerImparfait1)	FullyM	00:04
5.2.1.2. Verbes en -ARE à radical non palatalisant : «j'aimais» (52110-RentrerImparfait1)	HérémenceF	00:03
5.2.1.3. Verbes en -ARE à radical palatalisant : «je mangeais» (52120-MangerImparfait1)	IsérablesF	00:03
5.2.1.4. eram, *seram et stabam : «j'étais» (52130-EtreImparfait1)	HérémenceM	00:03
5.2.2.1. Verbes en - are à radical non palatalisant ; forme interrogative (question partielle) : «Que t'achetais-tu ?» (52210-AcheterImparfait2)	IsérablesM	00:02
5.2.2.2. Verbes en -are à radical palatalisant ; forme interrogative (question partielle) : «Que mangeais-tu ?» (52120-MangerImparfait1)	La Chapelle-d'AbondanceF	00:03
5.2.2.3. Verbes en -ēre ; subordonnée temporelle : «quand tu avais» (52230-AvoirImparfait2)	La Chapelle-d'AbondanceM	00:01
5.2.3.1. Verbes en - are à radical non palatalisant : «il/elle/on achetait» (52310-AcheterImparfait3)	LensF	00:07
5.2.3.2. Verbes en -are à radical palatalisant : «il/elle/on mangeait» (52320-MangerImparfait3)	Les MarécottesF	00:04
5.2.3.3. erat, *serat et stabat : «il/elle/on était» (52330-EtreImparfait3)	Les MarécottesM	00:03
5.2.3.6. «il/elle voulait» (52350-VouloirImparfait3)	LiddesF	00:04
5.2.3.7. «il/elle/on avait ; il y avait» (52360-AvoirImparfait3)	LiddesM	00:10
5.2.3.8. «il/elle venait/tenait » (52370-VenirTenirImparfait3)	MontanaF	00:04
5.2.4.1. Verbes en -are : morphème de l'imparfait et désinence (52410-AREImparfait4)	MontanaM	00:02

Suite du tableau 2 : Références des énoncés sélectionnés lors du carottage de l'ALAVAL

5.2.4.2. «nous étions» (52420-EtreImparfait4)	LourtierF	00:04
5.2.4.3. «nous avions» (52430-AvoirImparfait4)	LourtierM	00:03
5.2.4.4. «nous tenions» (52442-TenirImparfait4)	NendazF	00:03
5.2.4.5. «nous faisions» (52443-FaireImparfait4)	OrsièresF	00:07
5.2.4.6. «nous mettions» (52445-MettreImparfait4)	NendazM	00:02
5.2.5.1. Verbes en -are à radical non palatalisant : «vous achetiez» (52510-AcheterImparfait5Des)	St-JeanF	00:03
5.2.5.2. Verbes en -are à radical palatalisant : «vous mangiez» (52520-MangerImparfait5)	SavièseF	00:04
5.2.5.3. Verbes en -ire : «vous veniez» (52530-VenirImparfait5)	NendazF	00:02
5.2.6.1. «ils/elles achetaient» (52602-AcheterImparfait6)	OrsièresM	00:04
5.2.6.2. «ils/elles mangeaient» (52604-MangerImparfait6)	SixtF	00:09
5.2.6.3. «ils/elles allaient» (52606-AllerImparfait6)	SavièseM	00:11
5.2.6.4. «ils/elles étaient» (52610-EtreImparfait6)	St-JeanM	00:18
5.2.6.5. «ils/elles avaient» (52620-AvoirImparfait6)	SixtM	00:12
5.3.1. «je viendrai / me souviendrai» : la morphologie de la 1re personne du singulier (53110-VenirFutur1)	TorgnonF	00:02
5.3.2. «tu verras » : la morphologie de la 2e personne du singulier (53210-VoirFutur2)	TroistorrentsF	00:04
5.3.3.1. « il viendra» : la morphologie de la 3e personne du singulier (53310-VenirFutur3)	Val-d'IlliezF	00:07
5.3.3.2. « elle verra» : la morphologie de la 3e personne du singulier (53320-VoirFutur3)	ArbazF	00:03
5.3.4. La morphologie de la 1re personne du pluriel (53410-Futur4Des)	BionazF	00:02
5.3.5. Les désinences du futur simple de la 3e personne du pluriel (53610-Futur6Des)	TorgnonM	00:03
5.3.6. «ils/elles iront», radicaux et désinences (53620-AllerFutur6)	Troistor-rentsM	00:02

Suite du tableau 2 : Références des énoncés sélectionnés lors du carottage de l'ALAVAL

6.1.1. «j'aimerais» : la morphologie de la 1re personne du singulier (61110-AimerCond1)	Val-d'IlliezM	00:03
6.2.1. «tu serais» : les radicaux du conditionnel d'«être» (61210-EtreCond2Rad)	ChamosonF	00:05
6.1.3.1. «il faudrait» (61310-FalloirCond3)	VouvryM1	00:02
6.1.3.2. «elle aurait» (61320-AvoirCond3)	ChalaisF	00:04
6.1.4. «ils verraient» : la morphologie de la 3e personne du pluriel (61600-Conditionnel6)	ContheyF	00:06
6.1.2.1. 1re personne du singulier : «que je me lève» (62110-LeverSubjPres1)	VouvryM2	00:03
6.1.2.2. 1re personne du singulier : «que je mange» (62120-MangerSubjPres1)	ArbazM	00:03
6.1.2.3. 1re personne du singulier : «que je te voie» (62130-VoirSubjPres1)	ÉvolèneF	00:02
6.2.2. 2e personne du singulier : «que tu te lèves » (62210-LeverSubjPres2)	BionazM	00:02
6.2.3. 3e personne du singulier : «qu'il se lève» (62310-LeverSubjPre3)	ChalaisM	00:03
6.2.4. 1re personne du pluriel : «que nous nous levions» (62410-LeverSubjPres4)	ChamosonM	00:02
6.2.5. Morphologie du subjonctif présent, 2e personne du pluriel (62520-SubjPres5Des)	ÉvolèneM	00:04
6.2.6. 3e personne du pluriel : «qu'ils se lèvent » (62610-LeverSubjPres6)	ContheyM	00:03
6.2.7.1. Mode et temps après un verbe de la volonté au présent (62710-Vouloir)	FullyM	00:06
6.2.7.2. Mode et temps après un verbe de la volonté au passé (62720-Vouloir)	HérémenceF	00:03
6.2.7.3. Mode et temps après un verbe de la volonté au passé (62720-Vouloir)	HérémenceM	00:03
6.2.1. Énoncé hypothétique du présent/futur : degré élevé de probabilité (63100-VenirPropHyp (hypothèse très probable)	FullyF	00:03

Suite du tableau 2 : Références des énoncés sélectionnés lors du carottage de l'ALAVAL

6.2.2. Énoncé hypothétique du présent/futur : degré bas de probabilité (63200-VoirPropHyp (hypothèse peu probable))	IsérablesF	00:03
6.3.3. Énoncé hypothétique du passé : l'irréalité (63300-HypothetiqueIrreelle)	La Chapelle-d'AbondanceF	00:03
7.2. «tourner» comme verbe auxiliaire (retour et répétition) (71200-«tourner»)	LensF	00:06
7.3. L'inversion spontanée (71300-Inversion spontanee)	IsérablesM	00:04
7.4. La question partielle avec «que» (71320-QueQuestionPartielle)	La Chapelle-d'AbondanceM	00:01

II Requêtes utilisées lors de la première fouille du corpus ALIPE

Requête principale : liaisons codées comme Z (« erreur hors contexte »; cf Liégeois, 2014 : 183) :

- `combo +s"[^ Syntctx=Z expecCons=* realCons=* obliOpt=*]" *.cha @ +u`

Il existe une forme de redondance dans le système de codage de l'ALIPE. Une autre requête générale possible est :

- `combo +s"[^ Syntctx=* expecCons=* realCons=* obliOpt=2]" *.cha @ +u`

Requêtes secondaires :

- `combo +s"[^ Syntctx=* expecCons=* realCons=k obliOpt=*]" *.cha @ +u`[38] : pas de résultats
- `combo +s"[^ Syntctx=* expecCons=* realCons=p obliOpt=*]" *.cha @ +u`[38] : pas de pataquès
- `combo +s"[^ Syntctx=* expecCons=* realCons=r obliOpt=*]" *.cha @ +u`[38] : pas de pataquès
- `combo +s"[^ Syntctx=* expecCons=o realCons=z obliOpt=*]" *.cha @ +u`
- `combo +s"[^ Syntctx=* expecCons=o realCons=t obliOpt=*]" *.cha @ +u`
- `combo +s"[^ Syntctx=* expecCons=o realCons=n obliOpt=*]" *.cha @ +u`
- `combo +s"[^ Syntctx=* expecCons=o realCons=l obliOpt=*]" *.cha @ +u`
- `combo +s"[^ Syntctx=* expecCons=z realCons=t obliOpt=*]" *.cha @ +u`
- `combo +s"[^ Syntctx=* expecCons=z realCons=l obliOpt=*]" *.cha @ +u`
- `combo +s"[^ Syntctx=* expecCons=z realCons=n obliOpt=*]" *.cha @ +u`
- `combo +s"[^ Syntctx=* expecCons=t realCons=z obliOpt=*]" *.cha @ +u`
- `combo +s"[^ Syntctx=* expecCons=t realCons=n obliOpt=*]" *.cha @ +u`
- `combo +s"[^ Syntctx=* expecCons=t realCons=l obliOpt=*]" *.cha @ +u` : pas de résultats
- `combo +s"[^ Syntctx=* expecCons=n realCons=z obliOpt=*]" *.cha @ +u`
- `combo +s"[^ Syntctx=* expecCons=n realCons=t obliOpt=*]" *.cha @ +u` : pas de résultats
- `combo +s"[^ Syntctx=* expecCons=n realCons=l obliOpt=*]" *.cha @ +u`

[38] Ces requêtes sont plus larges que ce que je cherchais, puisqu'elles mènent à toutes les consonnes et pataquès réalisés en [(*)k], [(*)p] et [(*)r]. Ces trois consonnes étant rares ou absentes dans les corpus actuels, cela m'a permis de gagner un certain temps (voir requêtes suivantes pour les autres consonnes).

- `combo +s"[^ Syntctx=* expecCons=l realCons=* obliOpt=*]" *.cha @ +u`[39] : pas de résultats
- `combo +s"[^ Syntctx=* expecCons=r realCons=z obliOpt=*]" *.cha @ +u` : pas de résultats
- `combo +s"[^ Syntctx=* expecCons=r realCons=t obliOpt=*]" *.cha @ +u` : pas de résultats
- `combo +s"[^ Syntctx=* expecCons=r realCons=n obliOpt=*]" *.cha @ +u`
- `combo +s"[^ Syntctx=* expecCons=r realCons=l obliOpt=*]" *.cha @ +u` : pas de résultats
- `combo +s"[^ Syntctx=* expecCons=p realCons=z obliOpt=*]" *.cha @ +u` : pas de résultats
- `combo +s"[^ Syntctx=* expecCons=p realCons=t obliOpt=*]" *.cha @ +u` : pas de résultats
- `combo +s"[^ Syntctx=* expecCons=p realCons=n obliOpt=*]" *.cha @ +u` : pas de résultats
- `combo +s"[^ Syntctx=* expecCons=p realCons=l obliOpt=*]" *.cha @ +u` : pas de résultats
- `combo +s"[^ Syntctx=* expecCons=k realCons=* obliOpt=*]" *.cha @ +u`[40] : pas de résultats

[39] Requête générale afin de vérifier qu'aucun sandhi externe en [l] n'ait été codé comme liaison.

[40] Cette requête est volontairement large car les grands corpus oraux actuels ne présentent habituellement pas de liaison en /k/.

III Autres résultats aux requêtes utilisées lors de la première fouille du corpus ALIPE

- agglutinations de la consonne de liaison sur le M2 apparaissant dans un tour de parole séparé du tour de parole contenant le M1 :

(320) *FAT : et lui c'est [^ Syntctx = H expecCons = t realCons = o obliOpt = 0] un gros +...
%add : à CHI
*CHI : + + [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] our(s) !
(ALIPE, ali-prune-071120-1)

(321) *FAT : ouais mais là c'est des +...
%add : à CHI
*CHI : + + [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] étoiles !
(ALIPE, ali-prune-071120-1)

(322) *FAT : les petits +...
%add : à CHI
*CHI : + + [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] enfants.
(ALIPE, ali-prune-071120-1)

(323) *FAT : trois +...
%add : à CHI
*CHI : + + [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] oranges !
(ALIPE, ali-prune-071120-1)

(324) *FAT : non des gros +...
%add : à CHI
*CHI : + + [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] arbres.
(ALIPE, ali-prune-071124)

(325) *FAT : par contre il a des +... %add : à CHI
*CHI : + + [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] antennes.
(ALIPE, ali-prune-071124)

(326) *FAT : des ?
%add : à CHI
*CHI : +, [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] arrêtes.
(ALIPE, ali-prune-091127-1)

- des agglutinations de la consonne de liaison sur le M2 sans lien avec le tour de parole précédent :

(327) *CHI : xxx pa(r)ce que i(l) [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] en reste.
*CHI : xxx i(l) i(l) [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] en reste xxx +.. ?
(ALIPE, ali-baptiste-110720)

- occurrences que j'analyserais autrement :

(328) *FAT : hum moi j(e) voulais des [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] haricots mais [^ Syntctx = J expecCons = z realCons = o obliOpt = 0] i(l) y en [^ Syntctx = B expecCons = n realCons = n obliOpt = 1] avait pas.
(ALIPE, ali-prune-071120-2)

(329) *FAT : un [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] hibou.
%add : à CHI
*CHI : xx un [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] hibou pourquoi ?
*FAT : ben un [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] hibou.
%add : à CHI
(ALIPE, ali-prune-071122)

(330) *CHI : je peux chercher [^ Syntctx = I expecCons = r realCons = n obliOpt = 0] autre mais je peux.
(ALIPE, ali-salome-081121)

- exemples que je considère comme des pataquès :

(331) *CHI : +, eu de [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] autocollants (.) xxx demandé.
(ALIPE, ali-prune-071120-2)

(332) *MOT : <non> [<] non le dauphin le dauphin bleu ne remue pas des [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] yeux parc(e) que euh leurs [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] yeux [^ Syntctx = F expecCons = z realCons = o obliOpt = 0] i(ls) sont faits [^ Syntctx = I expecCons = z realCons = o obliOpt = 0] avec des [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] autocollants.

%add : à CHI
*CHI : un [^ Syntctx = A expecCons = n realCons = z obliOpt = 1][41] autocollant.
*MOT : hum.
%add : à CHI
*MOT : ce sont des [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] autocollants (.) qui sont collés (.) d'un côté et d(e) l'autre <et qui font> [>] +...
%add : à CHI
(ALIPE, ali-salome-081121)

(333) *CHI : i(l) y a <pas> [= ? p(l)ein] de [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] oiseaux.
(ALIPE, ali-baptiste-110713)

(334) *CHI : xx qui [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] ont pas +...
(ALIPE, ali-prune-071120-2)

(335) *CHI : +, des une [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] étoile deux [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] ét– +//.
(ALIPE, ali-prune-071120-1)

(336) *CHI : <attends> [<] attends faut mett(re) faut manger le [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] os comme ça.
*FAT : ah d'accord.
%add : à CHI
*FAT : le [^ Syntctx = Z expecCons = o realCons = n obliOpt = 2] os ?
%add : à CHI
(ALIPE, ali-prune-071120-1)

(337) *CHI : j– j– m– mon papa i(l) m'a dit que faudrait faudrait monter sur un [^ Syntctx = A expecCons = n realCons = n obliOpt = 1] hippopotame après sur un [^ Syntctx = A expecCons = n realCons = z obliOpt = 1] avion d'accord ?
*FAT : ah bon ?
%add : à CHI
*CHI : ben oui.
*FAT : mais faut pas monter sur un [^ Syntctx = A expecCons = n realCons = n obliOpt = 1] avion +...
%add : à CHI
(ALIPE, ali-prune-071121-2)

[41] Pour *autocollant*, Liégeois (2014 : 390) note que « ce Mot2 a tendance à être employé au pluriel ».

(338) *FAT : et lui c'est qui ?
%add : à CHI
*CHI : un [^ Syntctx = A expecCons = n realCons = n obliOpt = 1] avion !
*FAT : et s'i(l) y en [^ Syntctx = B expecCons = n realCons = n obliOpt = 1] avait [^ Syntctx = G expecCons = t realCons = o obliOpt = 0] un gros ça s(e)rait [^ Syntctx = H expecCons = t realCons = o obliOpt = 0] un +...
%add : à CHI
*CHI : un gros [^ Syntctx = E expecCons = z realCons = z obliOpt = 0] avion.
*FAT : un gros [^ Syntctx = E expecCons = z realCons = z obliOpt = 0] avion.
%add : à CHI
*CHI : et ça s(e)rait [^ Syntctx = I expecCons = t realCons = o obliOpt = 0] un petit [^ Syntctx = E expecCons = t realCons = z obliOpt = 0] avion.
%add : à CHI
(ALIPE, ali-prune-071124)

(339) *CHI : un [^ Syntctx = A expecCons = n realCons = z obliOpt = 1] oignon.
(ALIPE, ali-prune-071122)

(340) *CHI : ah je sais j'aime ça les [^ Syntctx = A expecCons = z realCons = n obliOpt = 0] hamburgers.
*FAT : ah oui ?
%add : à CHI
*CHI : tu peux pas te relever ?
*FAT : c'est bon les [^ Syntctx = A expecCons = z realCons = m obliOpt = 0] hamburgers hein ?
%add : à CHI
(ALIPE, ali-salome-090710)

(341) *CHI : mais je décolle les [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] animaux mais je les décolle.
*CHI : je décolle pa(r)c(e) que +...
*MOT : parc(e) que quoi ?
%add : à CHI
*CHI : r(e)ga(r)de j'ai un [^ Syntctx = A expecCons = n realCons = z obliOpt = 1] animaux aux mains sur ma main.
*CHI : <c'est> [= ? j'ai] un [^ Syntctx = A expecCons = n realCons = z obliOpt = 1] animaux sur ma main t(u) as vu ?
(ALIPE, ali-salome-081119)

(342) *CHI : et voulez une [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] assiette une assiette comme ça ?
[...]
*CHI : donnez-moi xxx[42].
*FAT : vot(re) quoi ?
%add : à CHI
[...]
*FAT : ah vot(re) [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] assiette ?
%add : à CHI
[...]
*CHI : donnez-moi vos [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] assiettes xxx du sel !
(ALIPE, ali-prune-071120-2)

(343) *CHI : ben vous voulez vot(re)[43] assiette ?
*FAT : euh quoi ?
%add : à CHI
*CHI : vous voulez vot(re) [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] assiette je vous dis !
(ALIPE, ali-prune-071122)

(344) *CHI : moi j(e) suis la maman et ça en fait c'était tout not(re) [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] enfant !
(ALIPE, ali-prune-071120-2)

(345) *CHI : c'est not(re) [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] enfant ça d'accord ?
(ALIPE, ali-prune-071121-2)

(346) *CHI : j(e) vais < chercher > [= ? r(e)chercher] mon [^ Syntctx = A expecCons = n realCons = z obliOpt = 1] enfant.
*CHI : +, les [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] enfants qui sont là-bas.
*FAT : d'accord.
%add : à CHI
*CHI : et après je reviens.
*FAT : hop là.

[42] Le père croit comprendre *votre* [*z] *assiette*.
[43] On entend déjà un pataquès en [*z] qui n'a pas été annoté.

%add : à CHI
*CHI : oh mon [^ Syntctx = A expecCons = n realCons = z obliOpt = 1] enfant !
(ALIPE, ali-prune-071121-2)

(347) *MOT : t(u) es le petit [^ Syntctx = E expecCons = t realCons = t obliOpt = 0] enfant de maman.
%add : à CHI
*CHI : non j(e) suis le grand [^ Syntctx = E expecCons = t realCons = n obliOpt = 0] enfant de maman !
*MOT : bravo !
%add : à CHI
*FAT : tu es la quoi ?
%add : à CHI
*MOT : t(u) as [^ Syntctx = G expecCons = z realCons = o obliOpt = 0] entendu ?
%add : à FAT
*FAT : non.
%add : à MOT
*MOT : je suis le grand [^ Syntctx = E expecCons = t realCons = n obliOpt = 0] enfant de maman.
%add : à FAT
*CHI : maman +...
*MOT : bravo c'est bien Prune.
%add : à CHI
*MOT : tout [^ Syntctx = C expecCons = t realCons = t obliOpt = 1] à l'heure elle a dit tous les [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] enfants aussi.
*CHI : xx i(l) y en [^ Syntctx = B expecCons = n realCons = n obliOpt = 1] avait comme +...
*CHI : +, i(l) y en [^ Syntctx = B expecCons = n realCons = n obliOpt = 1] avait comme ça maman ?
*MOT : < deux > [>] !
%add : à CHI
*FAT : < de quoi > [<] ?
%add : à CHI
*CHI : deux ?
*MOT : deux quoi ?
*CHI : ben deux [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] enfants.

*CHI : i(l) y en [^ Syntctx = B expecCons = n realCons = n obliOpt = 1] avait +...
*CHI : +, qu'un ?
*CHI : +, i(l) y en [^ Syntctx = B expecCons = n realCons = n obliOpt = 1] avait qu'un [^ Syntctx = P expecCons = n realCons = z obliOpt = 0] enfant maman ?
*MOT : <oh Prune> [>] +...
*CHI : <xxx> [<] xxx <xxx> [>].
*FAT : <et alors> [<] +...
*MOT : +, tu fais plaisir à ton papa.
(ALIPE, ali-prune-071121-1)

(348) *CHI : j'ai eu peur j'ai eu peur que maman elle me souhaite bon [^ Syntctx = P expecCons = n realCons = n obliOpt = 1] anniversaire ma p(e)tite chérie Salomé Patronyme-de-chi !
*MOT : ah oui +...
%add : à CHI
*MOT : pa(r)c(e) que t(u) as peur que j(e) te souhaite [^ Syntctx = I expecCons = z realCons = o obliOpt = 0] un bon [^ Syntctx = P expecCons = n realCons = n obliOpt = 1] anniversaire toi ?
%add : à CHI
*FAT : c'est pas tous les jours ton [^ Syntctx = A expecCons = n realCons = n obliOpt = 1] anniversaire hein.
%add : à CHI
*CHI : si c'est tous les jours mon p(e)tit [^ Syntctx = E expecCons = t realCons = n obliOpt = 0] anniversaire à m– à mon [^ Syntctx = A expecCons = n realCons = n obliOpt = 1] anni– à mon xxx.
(ALIPE, ali-salome-090713)

(349) *FAT : ben si c'est qui eux ?
%add : à CHI
*CHI : oh j(e) sais pas !
*FAT : ben !
%add : à CHI
*CHI : j(e) sais pas !
*FAT : c'est des [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] ours euh !
%add : à CHI
*CHI : et oui !
*FAT : et lui il est comment ?

%add : à CHI
*CHI : petit.
*FAT : c'est [^ Syntctx = H expecCons = t realCons = o obliOpt = 0] un p(e)tit quoi alors ?
%add : à CHI
*CHI : un p(e)tit [^ Syntctx = E expecCons = t realCons = z obliOpt = 0] ours.
*FAT : et lui c'est [^ Syntctx = H expecCons = t realCons = o obliOpt = 0] un gros +...
%add : à CHI
*CHI : + + [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] our(s) !
*FAT : et lui c'est [^ Syntctx = H expecCons = t realCons = o obliOpt = 0] un grand +...
%add : à CHI
*CHI : [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] ours !
*FAT : un grand quoi ?
%add : à CHI
*CHI : un grand [^ Syntctx = E expecCons = t realCons = z obliOpt = 0] ours euh !
(ALIPE, ali-prune-071120-1)

(350) *FAT : et imagine i(l) y aurait [^ Syntctx = G expecCons = t realCons = o obliOpt = 0] un < grand [^ Syntctx = E expecCons = t realCons = t obliOpt = 0] ours > [>] derrière toi.
%add : à CHI
*CHI : < non > [<].
*FAT : qu'est-c(e) que tu lui dirais ?
%add : à CHI
*CHI : vas[44]-t'en !
*FAT : vas-t'en qui ?
%add : à CHI
*CHI : vas-t'en l'ours !
*CHI : file de not(re) maison !
*CHI : va par la f(e)nêtre !
*FAT : et si c'était [^ Syntctx = H expecCons = t realCons = o obliOpt = 0] un petit ?
%add : à CHI

[44] [sic]

*CHI : ouais.
*FAT : tu lui dirais quoi ?
%add : à CHI
*CHI : vas-t'en petit.
*FAT : petit quoi ?
%add : à CHI
*CHI : petit xxx petit [^ Syntctx = E expecCons = t realCons = z obliOpt = 0] ours.
*CHI : petit xxx.
*CHI : xxx petit [^ Syntctx = E expecCons = t realCons = z obliOpt = 0] ours.
(ALIPE, ali-prune-080714-2)

(351) *CHI : < un petit [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] hibou > [= ! chuchote].
*FAT : ben dis-le plus fort j'entends pas.
%add : à CHI
*CHI : un petit [^ Syntctx = Z expecCons = o realCons = z obliOpt = 2] hibou.
(ALIPE, ali-prune-071124)

(352) *CHI : +, un grand [^ Syntctx = E expecCons = t realCons = z obliOpt = 0] arbre.
(ALIPE, ali-prune-071120-1)

(353) *FAT : qu'est-c(e) qu'un [^ Syntctx = A expecCons = n realCons = n obliOpt = 1] animal qui a quat(re) pattes et qui fait hi-han@o ?
%add : à CHI
*CHI : < un [^ Syntctx = A expecCons = n realCons = n obliOpt = 1] âne > [= ! crie] !
*FAT : et si i(l) y en [^ Syntctx = B expecCons = n realCons = n obliOpt = 1] a deux ?
%add : à CHI
*CHI : deux [^ Syntctx = A expecCons = z realCons = z obliOpt = 1] ânes.
*FAT : et si c'est [^ Syntctx = H expecCons = t realCons = o obliOpt = 0] un petit ?
%add : à CHI
*CHI : bébé.
*FAT : non si c'est [^ Syntctx = H expecCons = t realCons = o obliOpt = 0] un un [^ Syntctx = A expecCons = n realCons = n obliOpt = 1] âne petit ça fait [^ Syntctx = I expecCons = t realCons = o obliOpt = 0] un petit + ...
%add : à CHI
*CHI : + + bébé.

*FAT : mais non ça fait [^ Syntctx = I expecCons = t realCons = o obliOpt = 0] un petit quoi ?
%add : à CHI
*CHI : un petit 0 [= ! rit].
*CHI : un petit [^ Syntctx = E expecCons = t realCons = z obliOpt = 0] âne.
*FAT : non +...
%add : à CHI
*FAT : et si c'est [^ Syntctx = H expecCons = t realCons = o obliOpt = 0] un grand ?
%add : à CHI
*CHI : un une maman !
*FAT : mais non.
%add : à CHI
*FAT : ça fait [^ Syntctx = I expecCons = t realCons = o obliOpt = 0] un [^ Syntctx = A expecCons = n realCons = n obliOpt = 1] âne grand ça fait [^ Syntctx = I expecCons = t realCons = o obliOpt = 0] un grand +...
%add : à CHI
*CHI : 0 [= ! rit] <j(e) sais pas> [= ! rit].
*FAT : ben si !
%add : à CHI
*CHI : <j(e) sais pas> [= ! rit].
*CHI : <un p(e)tit bébé> [= ! rit].
*FAT : mais non !
%add : à CHI
*CHI : xxx un p(e)tit [^ Syntctx = E expecCons = t realCons = z obliOpt = 0] âne !
*FAT : non un grand !
%add : à CHI
*CHI : un grand.
*FAT : un grand quoi ?
%add : à CHI
*CHI : un grand [^ Syntctx = E expecCons = t realCons = z obliOpt = 0] âne.
(ALIPE, ali-prune-080714-2)

(354) *CHI : un p(e)tit [^ Syntctx = E expecCons = t realCons = n obliOpt = 0] oeuf dedans pic@o.
(ALIPE, ali-salome-090716)

IV Requêtes utilisées lors de la première fouille du corpus PFC

Requête principale : liaisons codées épenthétiques :

- `@liaison_presence 4`

Requêtes secondaires pour la fouille des données de lecture du texte :

- `@type "t" @cg_imm a`
- `@type "t" @liaison_consonne (z | n | r) @cg_imm grand`
- `@type "t" @cg_imm ils`
- `@type "t" @cg_imm mène`
- `@type "t" @cg_imm quatrième`
- `@type "t" @liaison_consonne (z | t) @cg_imm son`
- `@type "t" @liaison_consonne (z | t | n) @cg_imm provoquer` : pas de résultats

Requêtes secondaires pour la fouille des conversations :

- `@type ("g" | "l") @cg_imm a`
- `@cg_imm va` : pas de pataquès
- `@cg_imm sera` : pas de résultats
- `@cg_imm devra` : pas de résultats
- `@liaison_consonne (t | r | n) @cg_imm es`
- `@liaison_consonne (t | n | r) @cg_imm suis`
- `@liaison_consonne (z | r | n) @cg_imm est` : pas de pataquès
- `@liaison_consonne (t | r | n) @cg_imm sommes` : pas de résultats
- `@liaison_consonne (t | r | n) @cg_imm êtes` : pas de résultats
- `@liaison_consonne (z | r | n) @cg_imm sont` : pas de résultats
- `@liaison_consonne (z | r | n) @cg_imm était` : pas de résultats
- `@liaison_consonne (z | r | n) @cg_imm étaient` : pas de résultats
- `@liaison_consonne (t | r | n) @cg_imm étais` : pas de résultats
- `@cg_imm leur`
- `@liaison_consonne (z | t) @cg_imm un` : pas de pataquès
- `@liaison_consonne (z | t | n) @cg_imm quatre` : pas de résultats
- `@liaison_consonne (z | t | n) @cg_imm cinq` : pas de pataquès
- `@liaison_consonne (z | n) @cg_imm sept` : pas de résultats
- `@liaison_consonne (z | n) @cg_imm huit`
- `@liaison_consonne (z | t | n) @cg_imm neuf` : pas de résultats

- `@liaison_consonne (z | n) @cg_imm vingt`
- `@liaison_consonne (z | n) @cg_imm trente` : pas de résultats
- `@liaison_consonne (z | n) @cg_imm quarante` : pas de résultats
- `@liaison_consonne (z | n) @cg_imm cinquante` : pas de résultats
- `@liaison_consonne (z | n) @cg_imm soixante` : pas de résultats
- `@liaison_consonne (z | n) @cg_imm cent`
- `@liaison_consonne (z | t | n) @cg_imm mille` : pas de résultats
- `@liaison_consonne (z) @cg_imm qui`
- `@type ("g" | "l") @liaison_consonne (z | n | r) @cg_imm grand`
- `@liaison_consonne (t | n | r) @cg_imm gros` : pas de résultats
- `@cg_imm moi` : pas de résultats
- `@cg_imm toi` : pas de résultats
- `@cg_imm lui` : pas de résultats
- `@liaison_consonne (z | t | n | r) @cg_imm trop` : pas de résultats
- `@liaison_consonne (z | t | n | r) @cg_imm si` : pas de résultats

V Liaisons codées comme épenthétiques dans le corpus PFC

TABLEAU 3 : Liaisons codées comme épenthétiques dans le corpus PFC

locuteur	**enquete**	**transcription**	**pataques**
42acd1	Roanne	C : Place aux14z autres, dégagez, bien contents.’	0
svarb2	Nyon	RB : Et puis après20 il est11t arrivé en Suisse comme0414 euh, comme saisonnier, on disait.	0
31aoe1	Toulouse	La côte escarpée du Mont Saint-Pierre qui mène14t au village,	1
31alm1	Toulouse	qui mène14t au village,	1
bfasy1	Burkina Faso	BFASY1G : un11nVN ancien24z étudiant vient se présenter20 à toi, et puis, euh, ’moi je suis10 un tel, bon’ <E : Vous avez vu la fille tout à l’heure là ?>.	1
42amg1	Roanne	M : Pour les beaux jours, mais là quand14t on l’a assise dessus20 elle a eu un petit peu peur.	0
42amg1	Roanne	M : Alors je sais pas quand14t est-ce qu’elle va nous les faire ses petites dents.	0
42amg1	Roanne	M : Surtout quand14t on reste huit jours sans la voir euh.	0
42amg1	Roanne	M : Oh moi je trouve quand14t une enfant20 a déjà passé toute la journée à l’école.	0
69amg1	Lyon	MG : ils rénovent les, les, les11z immeubles de bas14t en haut, de bas11z en haut et,	1
42acl1	Roanne	C : Puis10 alors c’était20h euh une bavarde, (bruits de conversation) une bavarde née hein, quand14t elle est branchée sur quelque chose, alors là tu peux10 y aller.	0
ciasn1	Abidjan	CIASN1 : Euh, il m’a laissé la petite voiture, une 304, avec laquelle je me débrouillais pour me déplacer.	0
cqbdb1	Saguenay	Le Premier Ministre a14 en11nVN effet décidé de faire étape dans cette commune,	1

Suite du tableau 3 : Liaisons codées comme épenthétiques dans le corpus PFC

75csb1	Paris Centre ville	qui mène14t au village	1
13bma1	Aix-Marseille	La côte escarpée du Mont Saint-Pierre, qui mène24t au village. Non, qui mène au village.	1
sgarb1	Genève	qui mène14t au village	1
21acp1	Dijon	CP : des11z anciens24n esclaves, des11z anciens21z esclaves qui se sont10 enfuis. Et euh, du coup ils11z ont fait des petites communautés là-bas quoi.	1
92app1	Puteaux-Courbevoie	PP : en fait14 quand j'ai dit de, enfin j'ai pensé à être instit/, c'est plus parce que j'aimais bien, euh,	0
92app1	Puteaux-Courbevoie	PP : Ouais, donc, bon, je pourrais le faire, mais en fait14t se serait l'Adsem qui le ferait, quoi.	0
50ajm1	Brécey	JM : Alors il y a quelques gamins qui sont scolarisés quand/ quand14 c'est des familles de jeunes.	0
svapb1	Nyon	La côte escarpée du Mont Saint-Pierre qui mène14t au village	1
ciaak1	Abidjan	CIAAK1 : Je suis née le 24, quatrième mois quatre-vingt quatre.	0
cqbrt1	Saguenay	RT1 : Et puis quand14 mon frère a vendu son commerce en euh	0
cqbrt1	Saguenay	RT1 : quand14 c'/, quand14 c'est sa première euh,	0
cqbrt1	Saguenay	RT1 : quand14 c'/, quand14 c'est sa première euh,	0
42amg1	Roanne	Le vige, le village de Beaulieu est11t en grand14t émoi.	0
42amg1	Roanne	Qu'est-ce qui a donc valu à Beaulieu ce grand14t honneur ?	0
42ajl1	Roanne	J : Ben10 oui, ben je suis10 obligé d'y aller parce que le père euh. Il est, pas là, quand14t il est là il va jouer.	0
42ajl1	Roanne	J : Tous les jours à six14z heures20 il était20 au boulot, jusqu'à euh, dix14z heures du soir, hein.	0

Suite du tableau 3 : Liaisons codées comme épenthétiques dans le corpus PFC

42ajl1	Roanne	J : Tous les jours à six14z heures20 il était20 au boulot, jusqu'à euh, dix14z heures du soir, hein.	0
69acg1	Lyon	quatrième24z aux Jeux11z Olympiques de Berlin en mille neuf cent trente-six,	1
13amb1	Marseille Centre Ville	La côte escarpée du Mont Saint-Pierre qui mène14t au, qui mène au village,	1
scaog1	Neuchâtel	OG : comme0412 le1122 bar0411 euh d'hiver0414 <JB2 : Hum hum.>il tourne0422 beaucoup moins bien, <JB2 : ouais il est obligé de caser ses employés sur la terrasse XX.>Ouais.	0
scaog1	Neuchâtel	OG : Non pas le0112 délai-cad0452re, le0122 délai-cadre0421 il res0452te fixe0422 mais le1114 mon droit à toucher du pognon quoi. <JB2 : Ouais.>	0
rcark1	RCA Bangui	son14z us/ son11nVO usine de pâtes italiennes.	1
abafb1	Bejaia	La côte escarpée du Mont Saint-Pierre qui mène24 au village connaît des barrages chaque fois que les11z opposants de tous les bords manifestent leur colère.	1
92acd1	Puteaux-Courbevoie	La côte escarpée du Mont Saint-Pierre, qui mène14t au village,	1
sgajd1	Genève	JD : Et temps11z en, d/ dès qu'il fait nuit ben tous les euh14z alligators ben ils sortent. C'est10 à ce moment là qu'ils vont pêcher donc	1
21acl1	Dijon	Le village de Beaulieu est11t en grand14t émoi.	0
21acl1	Dijon	Qu'est-ce qui a donc valu à Beaulieu ce grand14t honneur ?	0
50alb1	Brécey	LB : Qui comptait soixante-quinze à quatre-vingt14z employés. E : Oui.	0

Suite du tableau 3 : Liaisons codées comme épenthétiques dans le corpus PFC

blajv1	Liège	BL : Compte Baladins euh quand moi j'y étais euh on11nVN était20 à plus je sais pas combien quoi hein maintenant, de nouveau ils sont s/ Pacs a dû mettre cent14z euros de sa poche là.	1
ciafk1	Abidjan	CIAFK1 : Compte0412 te1222nu de1114, de1132 l'attitude0412 belliqueuse0412 des gens de0112 la FESCI.	0
ciafk1	Abidjan	CIAFK1 : Et comme0412 là-bas, c'est l'année du bicente1312naire0414, on a voulu ici <E1 : Ah. >organiser, on a voulu ici organiser une0411 activité pour0412 voir041	0
ciafk1	Abidjan	CIAFK1 : que1134, que1131 ils considère0412nt comme0412 étant une1411, euh, faisant partie de1112 la rébellion	0
ciafk1	Abidjan	CIAFK1 : Donc0411, ils sont ve1212nus, ils nous ont délogés. Mais on a l'intention de1114, on va délocaliser les travaux.	0
ciafk1	Abidjan	CIAFK1 : le1132 Droit des Obligations, le1132 Droit des Affaire0413s, le1132 Droit Public0414, et, et cetera	0
ciafk1	Abidjan	CIAFK1 : Parce0352 que1424, l'idée c'était, voilà, on a hérité ça du Code0412 Civil0412 Napoléon, <E1 : Oui. >	0
ciafk1	Abidjan	CIAFK1 : Et, puis bon, euh, est-ce0312 que14224, ça a été bien reçu, bon.	0
ciafk1	Abidjan	CIAFK1 : Et donc0411, il0412 faut voir0414, ça fait s/ s/ c'est soixante0412-quatre0424, deux mille0412 quatre1422, ça fait quarante0411 ans <E1 : Hum hum. >que1132 ces lois là ont été prise0413s.	0
ciafk1	Abidjan	CIAFK1 : Et donc0411, il0412 faut voir0414, ça fait s/ s/ c'est soixante0412-quatre0424, deux mille0412 quatre1422, ça fait quarante0411 ans <E1 : Hum hum. >que1132 ces lois là ont été prise0413s.	0

Suite du tableau 3 : Liaisons codées comme épenthétiques dans le corpus PFC

ciafk1	Abidjan	CIAFK1 : La loi sur le1122 mariage0414, sur les successions,	0
ciafk1	Abidjan	CIAFK1 : Parce0322 que1424, la loi sur0412 la succession,	0
ciafk1	Abidjan	CIAFK1 : le1134, le1132 premier de1212gré, hein, euh, le1112 premier ordre0422 d'héritiers,	0
ciafk1	Abidjan	CIAFK1 : Mais au plan culturel0414, au plan culturel0414, ça ne1112 passe0412 pas.	0
ciafk1	Abidjan	CIAFK1 : Mais au plan culturel0414, au plan culturel0414, ça ne1112 passe0412 pas.	0
ciafk1	Abidjan	CIAFK1 : Parce0322 que1422 il y a(non codé) le1114, le1132 régime1411, euh, patriarcal0414, le1132 rég/ le1132 régime0412 matriarcal1414,	0
ciafk1	Abidjan	CIAFK1 : Parce0322 que1422 il y a(non codé) le1114, le1132 régime1411, euh, patriarcal0414, le1132 rég/ le1132 régime0412 matriarcal1414,	0
ciafk1	Abidjan	CIAFK1 : C'est le1112 ne1212veu qui doit hériter ou c'est l'oncle0124, ou c'est, c'est, c'est.	0
ciafk1	Abidjan	CIAFK1 : après, certaine0312ment après le1112 Colloque0414, on va essayer de1114, de1134, de1132 lancer cette1414, cette0411 étude0412 là	0
ciafk1	Abidjan	CIAFK1 : après, certaine0312ment après le1112 Colloque0414, on va essayer de1114, de1134, de1132 lancer cette1414, cette0411 étude0412 là	0
ciafk1	Abidjan	CIAFK1 : après, certaine0312ment après le1112 Colloque0414, on va essayer de1114, de1134, de1132 lancer cette1414, cette0411 étude0412 là	0
ciafk1	Abidjan	CIAFK1 : après, certaine0312ment après le1112 Colloque0414, on va essayer de1114, de1134, de1132 lancer cette1414, cette0411 étude0412 là	0
ciafk1	Abidjan	CIAFK1 : pour voir0414, au plan culturel0414, qu'est-ce0112 qu'il0412 faut faire0413.	0

Suite du tableau 3 : Liaisons codées comme épenthétiques dans le corpus PFC

ciafk1	Abidjan	CIAFK1 : pour voir0414, au plan culturel0414, qu'est-ce0112 qu'il0412 faut faire0413.	0
ciafk1	Abidjan	CIAFK1 : On a le1112 même0412 problème0412 par0412 rappor0411t à la justice1412 traditionnelle0414, parce0352 que1423.	0
ciafk1	Abidjan	CIAFK1 : C'est-à-dire qu'en fait0412 dans la Constitution ivoirienne0414, il y a(non codé) une0412 disposition qui,	0
ciafk1	Abidjan	CIAFK1 : on avait hérités de1114, de1132 la, de1132 la, de1134, de1134,	0
ciafk1	Abidjan	CIAFK1 : on avait hérités de1114, de1132 la, de1132 la, de1134, de1134,	0
21ash1	Dijon	SH : Il l'a su, il, il l'a su euh, quand14t il était presque majeur, alors, enfin.	0
13aog1	Marseille Centre Ville	OG1 : Alors le mec tu sais10 il14z a le, le, tu sais tu as le, le plat du dessus là le truc conique je sais plus comment ça s'appelle, tu sais10 il soulève il fait10h 'ah non, non me suis trompé c'est du poulet,'	0
21ash1	Dijon	SH : Tous les pou/, cours préparer à apprendre à lire et puis, deux14z élèves que j'ai menés au certificat d'études, et puis deux14z autres que j'ai préparés, enfin oui quoi c'est.	0
21ash1	Dijon	SH : Tous les pou/, cours préparer à apprendre à lire et puis, deux14z élèves que j'ai menés au certificat d'études, et puis deux14z autres que j'ai préparés, enfin oui quoi c'est.	0
91ael1	Brunoy	LE : Et puis finalement, en14t, en11nVN allant chez le boulanger, en11nVN arrivant à Brunoy, on voit10 une étiquette chiot10 à vendre.	0
75ccr2	Paris Centre ville	La côte escarpée du Mont Saint-Pierre qui mène14t au village	1
42ags1	Roanne	G : Hein, si c'est des deux14z étoiles, trois11z étoiles, quatre étoiles.	0

Suite du tableau 3 : Liaisons codées comme épenthétiques dans le corpus PFC

42ags1	Roanne	G : que si tu prends qu'un deux euh, un deux14z étoiles20 en France,	0
rcark1	RCA Bangui	RK1 : Les conseillers ne leurs14z aident pas10 à remédier à ce qu'ils11z ont dit auparavant.	1
92aaf3	Puteaux-Courbevoie	AF : Ca me paraît <E : Ca me paraît logique.>logique. Tu vas voir que quand14t c'est les miennes, euh, ce sera pas, ce sera l'inverse.	0
50aid1	Brécey	Le village de Beaulieu est11t un grand14nVN émoi	1
blasl1	Liège	La côte0414, c'est dégueulasse votre truc vous faites exprès de mettre (rires).	0
12ajp1	Rodez	quatrième24z aux jeux11z olympiques de Berlin en millle-cent-trente-six, et plus récemment, son11nVO usine de pâtes21z italiennes.	1
21ama1	Dijon	MB : Bon forcément celui qui y pense à midi, généralement c'est lui qui fait parce que celui qui arrive comme une fleur à six14z heure et demie : 'Tiens qu'est-ce qu'on va faire à manger ce soir?', ben c'est, c'est déjà un peu tard.	0
75xlv1	Aveyron-nais à Paris	Or, un gros détachement de police, comme on11nVO en11nVN a vu à Jonquière, et des vérifications d'identité risquent de provoquer24 une explosion.	1
50apb1	Brécey	Le Premier Ministre a14t en11nVN effet décidé de faire étape dans cette commune	1
50apb1	Brécey	qui mène14t au village connaît des barrages chaque fois	1
abahs1	Bejaia	Qu'est-ce qui14t a donc v/valu à Beaulieu ce grand11t honneur?	0
cqejg1	Trois-Rivières	JG : Ça fait, quatre ans que je suis14t avec les jeunes là,	0
cqejg1	Trois-Rivières	JG : avec un vrai14t investissement, là.	1
cqemc1	Trois-Rivières	MC : Je suis14t allé euh,	1

Suite du tableau 3 : Liaisons codées comme épenthétiques dans le corpus PFC

cqemc1	Trois-Rivières	MC : Je te dirais vraiment l'espagnol quand je suis14t arrivé là-bas c'est, c'est, c'est ce qui l'a fait développer là.	1
cqewd1	Trois-Rivières	WD : Pis10 après ça je suis14t allé g/ euh, Ninstitution.	1
cqewd1	Trois-Rivières	WD : Alors je suis14t allé faire mon cours, pis je suis rentré aux services20 après de,	1
cqewd1	Trois-Rivières	WD : Il y avait20 un problème à Nlieu, je suis14t allé ouvrir,	1
cqewd1	Trois-Rivières	WD : quand je suis14t entré en cinquante-neuf,	1
cqelc1	Trois-Rivières	LC : Euh, en Virginie ou euh, a/ aller20 à la plage. Je suis14t allée à Boston20h, à New York et pis euh. <EQ : Okay.>	0
cqelc1	Trois-Rivières	LC : Et pis moi ben j/, je suis14t allée à l'école, comme tout le monde.	1
cqejb1	Trois-Rivières	JB : à l'automne deux mille euh, dix à, en tant que modèle deux mille onze, ça va14t être des nouveautés, qui vont sortir qui vont10 être entièrement20 électriques.	0
cqell1	Trois-Rivières	LL : 'non, faut tout le temps que je change, je suis14t un vrai artiste'.	1
cqell1	Trois-Rivières	LL : (bruit de bouche) Je sais pas, ces temps-ci je suis14t assez facile à décevoir là, mais euh.	1
caaag3	Peace River	L : Là, là j'ai dit ça s'appelle 'kitchen0414 sink0424 soup0413' (ANG).	0
caaag3	Peace River	L : Là, là j'ai dit ça s'appelle 'kitchen0414 sink0424 soup0413' (ANG).	0
caalm1	Peace River	L : ben elle avait20 eu huit14z enfants, pis10 elle est morte avec le dernier bébé, elle était pas supposée de,	1
caarl1	Peace River	L : Là, là j'ai dit ça s'appelle 'kitchen0414 sink0424 soup0413' (ANG).	0

Suite du tableau 3 : Liaisons codées comme épenthétiques dans le corpus PFC

caarl1	Peace River	L : Là, là j'ai dit ça s'appelle 'kitchen0414 sink0424 soup0413' (ANG).	0
TOTAL			35

Abréviations

\# Frontière de mot
1 Première personne
2 Deuxième personne
3 Troisième personne
ABL Ablatif
ACC Accusatif

ADJ Adjectif
ADV Adverbe

ASS Associatif

AUD Auditeur

BRU Enquête réalisée par Ferdinand Brunot

ce Collection d'exemples
CHI *Child* (enfant)
cons *Consonantal* (consonantique)

DAT Datif
DEF Défini

DET Déterminant

DET:ART Article
DET:POS Déterminant possessif

F1 Premier formant
FAT *Father* (père)

F Féminin

fr. français

FUT Futur
GEN Génitif
IMP Impératif
IMPF Imparfait
IND Indicatif
INDEF Indéfini

KON Conjonction

LAKS Enquête réalisée par Bernard Laks
lat. latin

M Masculin

M1 Mot 1
M2 Mot 2
MOT *Mother* (mère)

NAM Nom propre

NOM Nominatif

NUM Numéral

OBL Oblique
PL Pluriel

PRO Pronom

PRO:DEM Pronom démonstratif

PRO:IND Pronom indéfini
PRO:PER Pronom personnel

PRP Préposition

PRP:det Préposition + déterminant
PRS Présent

s. siècle

SAP *Speech Act Participant*
SBJ Sujet
SG Singulier

SGR *Singer* (chanteur)
SN Syntagme nominal
SP Syntagme prépositionnel

SUP Superlatif

V Voyelle
VER Verbe

VER:impe Verbe conjugué à l'impératif présent
VER:infi Verbe à l'infinitif
VER:pper Participe passé
VER:ppre Participe présent

VIL Enquête de Villejuif
VILL Enquête de Villejuif

W1 *Word 1*
W2 *Word 2*

1, 2, 3, ABL, ACC, ASS, DAT, DEF, GEN, F, FUT, GEN, IMP, IMPF, IND, INDEF, M, NOM, OBL, PL, PRS, SAP, SBJ, SG et SUP correspondent à des abréviations communément utilisées dans les analyses morphosyntaxiques (gloses suivant les recommandations de Lehmann, 2004).
ADJ, ADV, DET:ART, DET:POS, KON, NAM, NUM, PRO, PRO:DEM, PRO:IND, PRO:PER, PRP, PRP:det, VER:impe, VER:infi, VER:pper et VER:ppre correspondent à des abréviations utilisées dans les étiquettes de Treetagger pour le français (étiquetage en parties du discours ; voir Stein, 2003 ; Stein & Schmid, 1995).

Acronymes

2 CV Deux chevaux

AF Ancien français

ALAVAL Atlas linguistique audiovisuel des dialectes francoprovençaux du Valais romand

ALIPE Acquisition de la liaison et interaction parents enfants

API Alphabet phonétique international

BIPO Base inter-régionale du patrimoine oral

BnF Bibliothèque nationale de France

CEMÉA Centres d'entraînement aux méthodes d'éducation active

CFPP2000 Corpus de français parlé parisien des années 2000

CHAT *Codes for the Human Analysis of Transcripts*

CHILDES *Child Language Data Exchange System*

CLAN *Computerized Language ANalysis*

CLARIN *Common Language Resources and Technology Infrastructure*

CMTRA Centre des musiques traditionnelles en Rhône-Alpes

COCOON COllections de COrpus Oraux Numériques

CORLI Corpus, langues, interactions

ELAN *EUDICO Linguistic Annotator*

ESLO Enquêtes sociolinguistiques à Orléans

ETAPE Évaluation en Traitement Automatique de la ParolE

FAMDT Fédération des associations de musiques et danses traditionnelles

FC Français contemporain

FJD Fonds Jean Dumas

FLE Français langue étrangère

FLS Français langue seconde

HPOL Hommes politiques

IFOP Institut français d'opinion publique

INA Institut national de l'audiovisuel

IPFC InterPhonologie du français contemporain

IPSE Indice de position socioéconomique

ISO Organisation internationale de normalisation

JD Jean Dumas

LED Langue en danger

LLL Laboratoire Ligérien de Linguistique

MOOC *Massive Open Online Course*

NCCFr *Nijmegen Corpus of Casual Speech*

NOTRe Nouvelle organisation territoriale de la République

OLAC *Open Language Archives Community*

ORTOLANG Plateforme d'Outils et de Ressources linguistiques pour un Traitement Optimisé de la LANGue française

OT *Optimality Theory*(Théorie de l'optimalité)

PDF *Portable Document Format*

PFC Phonologie du français contemporain

POS *Part of speech*

SAMPA Speech Assessment Methods Phonetic Alphabet

SNEP Syndicat national de l'édition phonographique

TGIR Très grande infrastructure de recherche

VLO *Virtual Language Observatory*

Index

Agglutination, 64
Aksak, 189

Candidat (OT), 24
Carottage, 130
Chanson
    Chanson de tradition orale, 178
    Chanson populaire, 178
    Chanson signée, 197
    Chanson traditionnelle, 178
Chants de mai, 187
Collection d'exemples, 122
Compétence, 251
Consonne
    Consonne de liaison, 22
    Consonne fixe, 23
    Consonnes équivalentes, 85
Corpus
    Corpus annoté, 127
    Corpus brut, 127
    Corpus clos, 126
    Corpus de référence, 127
    Corpus de suivi, 126
    Corpus fermé, 126
    Corpus ouvert, 127
    Corpus spécialisé, 127
Cuir, 90

Diction
    Diction directe, 27
    Diction indirecte, 27
Données
    Données primaires, 127
    Données secondaires, 127
    Données élicitées, 132
Déglutination, 64
Déterminant unipossessif, 42

Édulcoration, 196
Enchaînement, 26
Épithèse, 94
Erreur
    Erreur d'omission, 98
    Erreur par régularisation, 99
Études en temps apparent, 66
Études en temps réel, 66
Euphonie, 80
    ‹s› euphonique, 80
    ‹t› euphonique, 80

‹h›- aspirés, 37
‹h›- disjonctifs, 37
‹h›- muets, 37
Hiatus, 23
    Hiatus interne, 24
    Hiatus transitoire, 24

Liaison, 21
    Contexte de liaison, 21
    Liaison attestée, 48
    Liaison avec enchaînement, 26
    Liaison enchaînée, 26
    Liaison erratique, 47
    Liaison facultative, 46
    Liaison interdite, 46
    Liaison invariable, 46
    Liaison non attestée, 47
    Liaison non enchaînée, 27
    Liaison obligatoire, 46
    Liaison sans enchaînement, 27
    Liaison variable, 48
    Site possible de liaison, 21
    Site potentiel de liaison, 21
Liaison vocalique, 22
Liguème, 22

Mot
    Mot 1, 21
    Mot 2, 21
    Mot liaisonnant, 21
    Mot liaisonné, 21
Métanalyse, 64
Méthode boule de neige, 187

Norme
    Norme descriptive, 91
    Norme prescriptive, 91

Oxyton, 56

Paradoxe de l'informateur, 194
Paragoge, 94
    Phonème paragogique, 94
Paroxyton, 57
Pataquès, 97
    Pataquès en contexte de liaison, 100
    Pataquès hors contexte de liaison, 100
    Pataquès paradigmatique, 100
    Pataquès syntagmatique, 100
Performance (linguistique), 251
Performance (musicologie), 179
Proparoxyton, 57
Psilose, 63

Radical, 38
Règle de la liaison supposée, 85

Sandhi, 19
    Sandhi consonantique, 20
    Sandhi externe, 19
    Sandhi interne, 19
    Sandhi tonal, 20
    Sandhi vocalique, 20
Syllabe
    Pénultième, 57
    Antépénultième, 57
    Syllabe tonique, 57

Thème (morphologie), 38
    Espace thématique, 38
Tire, 136

Variation
    Variation diagénique, 119
    Variation diamésique, 77
    Variation diaphasique, 82
    Variation diastratique, 87
    Variation diatopique, 80
Variété
    Variété standard, 93
    Variétés américaines, 81

Variétés canadiennes, 81
Velours, 90
Verbe tensé, 41
Word
Word 1, 21
Word 2, 21

Crédits des illustrations

AGENCE DES MUSIQUES DES TERRITOIRES D'AUVERGNE (AMTA) & DESGRUGILLERS, Eric, [s. d.]. *Fiche 145 (JD_a060_29) [Fiche de chanson tapuscrite]*.

DUMAS, Jean, [s. d.(a)]. *Fiche 0140 (JD_a060_24) [Fiche de chanson manuscrite]*. Disponible à l'adresse : patrimoine-oral.org/seam/resource/directMedia/l7NM8hn6zn4TmNDWgXxVoo0Kh binaryFileId=6151&cid=1824.

DUMAS, Jean, [s. d.(b)]. *Fiche 0156 (JD_a061_01) [Fiche de chanson manuscrite]*. Disponible à l'adresse : patrimoine-oral.org/seam/resource/directMedia/l7NM8hn6zn4TmNDWgXxVokSza binaryFileId=6338&cid=1816.

DUMAS, Jean, [s. d.(c)]. *Fiche 0158 (JD_a061_03) [Fiche de chanson manuscrite]*. Disponible à l'adresse : patrimoine-oral.org/seam/resource/directMedia/l7NM8hn6zn4TmNDWgXxVohO8g 92?binaryFileId=6344&cid=1808.

DUMAS, Jean, [s. d.(d)]. *Fiche 0159 (JD_a061_04) [Fiche de chanson manuscrite]*. Disponible à l'adresse : patrimoine-oral.org/seam/resource/directMedia/Z_ 4FtudNgrCmZh-F03XtV9ErXo0Ij6Qp?binaryFileId=6347&cid=2350.

DUMAS, Jean, [s. d.(e)]. *Fiche 0163 (JD_a061_08) [Fiche de chanson manuscrite]*. Disponible à l'adresse : patrimoine-oral.org/seam/resource/directMedia/Z_ 4FtudNgrCmZh-F03XtVxUstNtgb8zv?binaryFileId=6429&cid=2558.

DUMAS, Jean, [s. d.(f)]. *Fiche 0168 (JD_a061_13) [Fiche de chanson manuscrite]*. Disponible à l'adresse : patrimoine-oral.org/seam/resource/directMedia/Z_ 4FtudNgrCmZh-F03XtVyTPDoV3Myc-?binaryFileId=6443&cid=2350.

DUMAS, Jean, [s. d.(g)]. *Fiche 0169 (JD_a061_14) [Fiche de chanson manuscrite]*. Disponible à l'adresse : patrimoine-oral.org/seam/resource/directMedia/Z_ 4FtudNgrCmZh-F03XtV83uC89H-LRl?binaryFileId=6446&cid=2350.

DUMAS, Jean, [s. d.(h)]. *Fiche 145 (JD_a060_29) [Fiche de chanson manuscrite]*.

DUMAS, Jean, 1960. *Séance d'enregistrement des dentellières par Thelma [Photographie]*.

DUMAS, Thelma, [s. d.(i)]. *Jean Dumas et sa 2 CV [Photographie]*.

GARDETTE, Pierre Mgr (éd.), 1950a. *1172. Les (garçons); les (oiseaux) [Carte]*. Paris; Lyon : CNRS; Institut de linguistique romane des Facultés Catholiques de Lyon. Disponible à l'adresse : https://www.ortolang.fr/market/corpora/ally/v1.

GARDETTE, Pierre Mgr (éd.), 1950b. *1174. Les (filles); les (oreilles) [Carte]*. Paris; Lyon : CNRS; Institut de linguistique romane des Facultés Catholiques de Lyon. Disponible à l'adresse : https://www.ortolang.fr/market/corpora/ally/v1.

GOOGLE BOOKS NGRAM VIEWER, 2019. *le hiatus,l'hiatus,de hiatus,d'hiatus [Requête]* [en ligne]. Consulté le 30/03/2021. Disponible à l'adresse: https://books.google.com/ngrams/graph?content=le+hiatus%2Cl%27hiatus%2Cde+hiatus%2Cd%27hiatus&year_start=1800&year_end=2019&corpus=fr-2019&smoothing=3.

INTERNATIONAL PHONETIC ASSOCIATION, 2018. *International Phonetic Alphabet Chart (IPA)*. Disponible à l'adresse: http://www.internationalphoneticassociation.org/content/ipa-chart.

JEANDILLOU, Jean-François, 2019. *Rappels précisonnels concernant les Comités de suivi de thèse [Mail]*.

Les Zenfants zéro déchet - ze mission [Illustration], [s. d.]. Disponible à l'adresse : https://www.thierrysouccar.com/bien-etre/livre/les-zenfants-zero-dechet-ze-mission-3482.

Ozongles [Photographie], [s. d.]. Disponible à l'adresse : https://www.pagesjaunes.fr/pros/56394659.

POUCHET, Laurent, 1914. *Laurent Pouchet à ses parents le 01-10-1914 depuis Vermelles [Lettre]*. Disponible à l'adresse : https://hdl.handle.net/11403/corpus14/v1.

Virginie Granouillet et Jean Dumas en 1960 [Photographie], [s. d.]. Disponible à l'adresse : https://www.leprogres.fr/sortir/2014/09/18/l-oralite-dans-le-velay-au-coeur-des-cahiers-de-la-haute-loire.

Bibliographie

Bibliographie scientifique

ABEILLÉ, Anne & GODARD, Danièle, 2007. Les relatives sans pronom relatif. In : *Le francais parlé au XXIème siècle: normes et variations dans les discours et en interaction*. Oxford : L'Harmattan. T. 2, pp. 37-60. Espaces discursifs (cf. p. 109).

ADDA-DECKER, Martine ; FOUGERON, Cécile ; GENDROT, Cédric ; DELAIS-ROUSSARIE, Elisabeth & LAMEL, Lori, 2012. La liaison dans la parole spontanée familière: une étude sur grand corpus. *Revue française de linguistique appliquée*. T. XVII, n° 1, pp. 113-128 (cf. p. 32-35, 46, 50, 147, 149, 150, 268, 286, 287).

ÅGREN, John, 1973. *Étude sur quelques liaisons facultatives dans le français de conversation radiophonique: fréquences et facteurs*. Kå-We Tryck, Uppsala : Acta Universitatis Upsaliensis. Studia Romanica Upsaliensia, n° 10 (cf. p. 41, 46, 50, 67).

AHMAD, Mohammad, 1993. *Vingt heures de français parlé... Aspects phonétiques de liaison*. Grenoble. Thèse de doctorat: Sciences du langage. Université Stendhal Grenoble III (cf. p. 29, 32, 34, 46, 49, 50).

AMMON, Ulrich ; DITTMAR, Norbert ; MATTHEIER, Klaus J. & TRUDGILL, Peter (éd.), 1987. *Sociolinguistics: An International Handbook of the Science of Language and Society/Soziolinguistik: Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft*. T. Volume 1. 2ème édition (2004). Berlin; New York : Walter de Gruyter (cf. p. 93).

AQUIEN, Michèle, 1993. *Dictionnaire de poétique*. Paris : Librairie Générale Française. Guides de la langue française (cf. p. 84, 85).

ARMSTRONG, Nigel & POOLEY, Tim, 2010. *Social and Linguistic Change in European French*. New York : Palgrave Macmillan (cf. p. 67, 70, 71, 90, 116).

AROM, Simha, 2004. L'aksak: Principes et typologie. *Cahiers d'ethnomusicologie (anciennement Cahiers de musiques traditionnelles)*. N° 17, pp. 11-48 (cf. p. 189).

ASHBY, William, 2003. *La liaison variable en français parlé tourangeau: une analyse en temps réel*. Tours, France (cf. p. 68, 70, 71).

BARRAS, Claude; GEOFFROIS, Edouard; WU, Zhibiao & LIBERMAN, Mark, 1998. Transcriber: a free tool for segmenting, labeling and transcribing speech. In: *Proceedings of the 1st International Conference on Language Resources and Evaluation 1998*. Granada, Spain (cit. on p. 144).

BARRECA, Giulia, 2015. *L'acquisition de la liaison chez des apprenants italophones: Des atouts d'un corpus de natifs pour l'étude de la liaison en français langue étrangère (FLE)*. Nanterre. Thèse de doctorat: Sciences du langage. Université Paris Ouest Nanterre La Défense, Università Cattolica del Sacro Cuore di Milano (cf. p. 39, 40, 50, 259, 260, 262, 264, 297).

BARRECA, Giulia ; CHRISTODOULIDES, George ; DURAND, Jacques ; LAKS, Bernard & LYCHE, Chantal, 2015. *La liaison de l'adjectif antéposé dans le français parlé*. Caen, France (cf. p. 25).

BAUCHE, Henri, 1920. *Le langage populaire: Grammaire, syntaxe et dictionnaire du français tel qu'on le parle dans le peuple de Paris avec tous les termes d'argot usuel*. Paris : Payot (cf. p. 36, 40, 65, 90).

BAUDE, Olivier ; BLANCHE-BENVENISTE, Claire ; CALAS, Marie-France ; CAPPEAU, Paul ; CORDEREIX, Pascal ; GOURY, Laurence ; JACOBSON, Michel ; DE LAMBERTERIE, Isabelle ; MARCHELLO-NIZIA, Christiane & MONDADA, Lorenza, 2006. *Corpus oraux: Guide des bonnes pratiques 2006*. Orléans; Paris : Presses Universitaires d'Orléans; CNRS Editions (cf. p. 127, 129, 133, 137, 190).

BAUDE, Olivier & DUGUA, Céline, 2015. Usages de la liaison dans le corpus des ESLOs: vers de nouveaux (z)ouvrages de référence? In : *La dia-variation en français actuel: Etudes sur corpus, approches croisées et ouvrages de référence*. G. Dostie & P. Hadermann. Bern; Berlin; Bruxelles; Frankfurt am Main; New York; Oxford; Wien : Peter Lang. T. 116, pp. 349-371. Sciences pour la communication (cf. p. 46, 69, 130).

BAYLEY, Robert ; CAMERON, Richard & LUCAS, Ceil, 2013. *The Oxford Handbook of Sociolinguistics*. New York : Oxford University Press (cf. p. 93).

BAYLON, Christian, 1996. *Sociolinguistique: société, langue et discours*. 3ème édition (2005). Paris : Armand Colin. Fac: Linguistique (cf. p. 91, 93).

BÉCHADE, Hervé D., 1992. *Phonétique et morphologie du français moderne et contemporain*. Paris : Presses universitaires de France. PUF fondamental (cf. p. 90).

BELLY, Marlène, 2015. Quand la chanson mène la danse: de Coirault à Guilcher. *Ethnologie française*. T. 45, n° 3, pp. 566-569 (cf. p. 185).

BERCHTOLD, Elisabeth ; BOERI, Gisèle ; BONVIN, Steve ; DIÉMOZ, Federica ; KRISTOL, Andres ; MAÎTRE, Raphaël ; MARQUIS, Chiara ; PANNATIER, Gisèle ; REUSSER-ELZINGRE, Aurélie & SAUZET, Maguelone, 2019a. *Atlas linguistique audiovisuel du francoprovençal valaisan: Commentaire linguistique des cartes*. Centre de dialectologie et d'étude du français régional de l'Université de Neuchâtel (cf. p. 214).

BERCHTOLD, Elisabeth ; BOERI, Gisèle ; BONVIN, Steve ; DIÉMOZ, Federica ; KRISTOL, Andres ; MAÎTRE, Raphaël ; MARQUIS, Chiara ; PANNATIER, Gisèle ; REUSSER-ELZINGRE, Aurélie & SAUZET, Maguelone, 2019b. *Atlas linguistique audiovisuel du francoprovençal valaisan: Questions de morphologie et syntaxe francoprovençales: Introduction*. Centre de dialectologie et d'étude du français régional de l'Université de Neuchâtel (cf. p. 31, 209, 214, 216).

BERGERON-MAGUIRE, Myriam, 2021. "'[D]es navir qui ne vienderons poien ta la pointe': quelques témoignages remarquables pour l'histoire de la liaison en français". *Revue de linguistique romane*. T. 85, pp. 75-99 (cf. p. 78).

BERT, Michel & GRINEVALD, Colette, 2010. Proposition de typologie des locuteurs de LED. *Faits de langues*. N° 35-36, pp. 117-132 (cf. p. 186).

BLAINEY, Darcie L., 2013. *First to come, last to go: Phonological change and resilience in Louisiana Regional French*. New Orleans. Thèse de doctorat. Tulane University (cit. on p. 169).

BLANCHE-BENVENISTE, Claire, 1997. *Approches de la langue parlée en français*. 2ème édition (2010). Paris : Ophrys. L'essentiel français (cf. p. 90, 105, 116, 203).

BLANCHE-BENVENISTE, Claire, 2004. Le singulier et le pluriel en français parlé contemporain. *Bulletin de la Société de linguistique de Paris*. T. XCIC, n° 1, pp. 129-154 (cf. p. 91).

BLANCHE-BENVENISTE, Claire, 2010. *Le français: usages de la langue parlée*. Leuven; Paris : Peeters. Les langues du monde (cf. p. 20, 42, 44, 47, 53, 54, 106, 108, 116, 117).

BLANCHET, Philippe, 2000. *Linguistique de terrain, méthode et théorie: Une approche ethno-sociolinguistique de la complexité*. 2ème édition (2012). Rennes : Presses Universitaires de Rennes (cf. p. 194).

BONAMI, Olivier & BOYÉ, Gilles, 2003a. La nature morphologique des allomorphies conditionnées: les formes de liaison des adjectifs en français. *Silexicales*. N° 3, pp. 39-48 (cf. p. 39, 43).

BONAMI, Olivier & BOYÉ, Gilles, 2003b. Supplétion et classes flexionnelles. *Langages*. N° 152, pp. 102-126 (cf. p. 38).

BOOIJ, Geert & DE JONG, Daan, 1987. The domain of liaison: theories and data. *Linguistics*. T. 25, n° 5, pp. 1005-1025 (cf. p. 39).

BORDAL, Guri, 2012. A phonological study of French spoken by multilingual speakers from Bangui, the capital of the Central African Republic. In : *Phonological Variation in French: Illustrations from three continents*. R. Gess, Ch. Lyche & T. Meisenburg. Amsterdam; Philadelphia : John Benjamins. T. 11, pp. 23-43. Studies in Language Variation (cf. p. 34, 82).

BORÉ, Catherine & ELALOUF, Marie-Laure, 2017. Deux étapes dans la construction de corpus scolaires: problèmes récurrents et perspectives nouvelles. *Corpus*. N° 16 (cf. p. 77).

BOUTIN, Béatrice Akissi & LYCHE, Chantal, 2014. Ce que nous apprennent des locuteurs francophones non-lecteurs sur la liaison. In : *La liaison: approches contemporaines*. Ch. Soum-Favaro, A. Coquillon & J.-P. Chevrot. Bern; Berlin; Bruxelles; Frankfurt am Main; New York; Oxford; Wien : Peter Lang. T. 110, pp. 283-310. Sciences pour la communication (cf. p. 81, 117, 118).

BRADLEY, Evan D, 2018. A comparison of the acoustic vowel spaces of speech and song. *Linguistic Research*. Vol. 35, no. 2, pp. 381–394 (cit. on p. 251).

BRANDÃO DE CARVALHO, Joaquim, 2018. Je ll'ai vu: Perception-driven allomorphic optimization of French l'. *Probus*. Vol. 30, no. 1, pp. 47–66 (cit. on pp. 109, 110).

BRUGMAN, Hennie & RUSSEL, Albert, 2004. Annotating Multi-media/Multi-modal resources with ELAN. In: *Proceedings of the 4th International Conference on Language Resources and Evaluation 2004*. Lisbon, Portugal: European Language Resources Association (ELRA) (cit. on p. 144).

BUBEN, Vladimir, 1935. *Influence de l'orthographe sur la prononciation du français moderne*. T. 1. Bratislava : Filosofická fakulta University Komenského (cf. p. 58).

BULOT, Thierry & BLANCHET, Philippe, 2013. *Une introduction à la sociolinguistique: pour l'étude des dynamiques de la langue française dans le monde*. Paris : Archives contemporaines (cf. p. 93).

BÜRKI, Audrey & LAGANARO, Marina, 2014. Représentation et encodage de la liaison: approches psycholinguistiques et neuropsycholinguistiques. In : *La liaison: approches contemporaines*. Ch. Soum-Favaro, A. Coquillon & J.-P. Chevrot. Bern; Berlin; Bruxelles; Frankfurt am Main; New York; Oxford; Wien : Peter Lang. T. 110, pp. 169-185. Sciences pour la communication (cf. p. 21).

BUSCAIL, Laurie & LYCHE, Chantal, 2016. French in Orne (Basse-Normandie): A speaker from Domfront. In : *Varieties of Spoken French*. S. Detey, J. Durand, B. Laks & Ch. Lyche. Oxford : Oxford University Press, pp. 159-168. Oxford Linguistics (cf. p. 95).

BYBEE, Joan, 2002. Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition*. T. 24, nº 2, pp. 215-221 (cf. p. 58).

BYBEE, Joan, 2005. La liaison: effets de fréquence et constructions. *Langages*. T. 2, nº 158, pp. 24-37 (cf. p. 40, 45, 58, 105).

CAPPEAU, Paul & GADET, Françoise, 2007a. Document 3: Où en sont les corpus sur les français parlés ? *Revue française de linguistique appliquée*. T. XII, nº 1, pp. 129-133 (cf. p. 128, 129).

CAPPEAU, Paul & GADET, Françoise, 2007b. L'exploitation sociolinguistique des grands corpus: Maître-mot et pierre philosophale. *Revue française de linguistique appliquée*. T. XII, nº 1, pp. 99-110 (cf. p. 124).

CAPPEAU, Paul & SEIJIDO, Magalie, 2005. *Les corpus oraux en français (inventaire 2005 v.1.0)*. DGLFLF (cf. p. 127).

CARTON, Fernand, 1974. *Introduction à la phonétique du français*. 2ème édition (1979). Mayenne : Bordas. Études (cf. p. 29, 30, 40).

CERISARA, Christophe; MELLA, Odile & FOHR, Dominique, 2009. JTrans, an open-source software for semi-automatic text-to-speech alignment. In : *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*. Brighton, United Kingdom (cf. p. 144).

CHARLES-DOMINIQUE, Luc, 2007. Patrice COIRAULT: Répertoire des chansons de tradition orale. III. Religion, crimes, divertissements. Paris, BnF, 2006. *Cahiers d'ethnomusicologie (anciennement Cahiers de musiques traditionnelles)*. Nº 20, pp. 330-334 (cf. p. 185).

CHAUVIN, Carole, 1999. *Comptines, formulettes et jeux enfantins dans les Alpes occidentales (région Rhône-Alpes, Suisse romande et Val d'Aoste): étude gestuelle, rythmique et verbale*. Grenoble. Thèse de doctorat. Université Stendhal Grenoble III (cf. p. 86).

CHAUVIN, Carole & COLLETTA, Jean-Marc, 2002. La gestualité dans les jeux chantés du folklore enfantin : description, transcription et analyse. In : *Sémiotiques non verbales et modèles de spatialité : Textes du congrès Sémio 2001*. G. Barrier & N. Pignier. Limoges : Presses Universitaires de Limoges, pp. 39-62 (cf. p. 86).

CHEN, Matthew Y., 2000. *Tone Sandhi: Patterns Across Chinese Dialects*. Cambridge: Cambridge University Press. Cambridge Studies in Linguistics, no. 92 (cit. on p. 20).

CHEVALIER, Catherine & FRÉCHETTE, Roseline, 2004. De m'amie à ma mie: étude diachronique des déterminants unipossessifs. In : *Actes du 8e Colloque des étudiantes et étudiants en Sciences du Langage*. Montréal, Canada (cf. p. 42, 43, 72).

CHEVROT, Jean-Pierre ; CHABANAL, Damien & DUGUA, Céline, 2007. Pour un modèle de l'acquisition des liaisons basé sur l'usage: trois études de cas. *Journal of French Language Studies*. T. 17, n° 1, pp. 103-128 (cf. p. 46).

CHEVROT, Jean-Pierre; DUGUA, Céline; HARNOIS-DELPIANO, Mylène; SICCARDI, Anne & SPINELLI, Elsa, 2013. Liaison acquisition: debates, critical issues, future research. *Language Sciences*. No. 39, pp. 83–94 (cit. on pp. 98, 99).

CHEVROT, Jean-Pierre & FAYOL, Michel, 2001. Acquisition of French Liaison and Related Child Errors. In : *Research on Child Language Acquisition: Proceedings of the 8th Conference of the International Association for the Study of Child Language*. San Sebastian, Spain : Cascadilla Press. T. 2, pp. 761-775 (cf. p. 21).

CHIGAREVSKAÏA, Nina Aleksandrovna, 1973. *Traité de phonétique française: cours théorique*. 3ème édition (1982). Moscou : Vysšaja škola (cf. p. 65).

CHOMSKY, Noam, 1965. *Aspects of the Theory of Syntax*. Cambridge (Massachusetts): The MIT Press. No. 11 (cit. on pp. 251, 252).

COQUILLON, Annelise & TURCSAN, Gabor, 2012. An overview of the phonological and phonetic properties of Southern French: Data from two Marseille surveys. In : *Phonological Variation in French: Illustrations from three continents*. R. Gess, Ch. Lyche & T. Meisenburg. Amsterdam; Philadelphia : John Benjamins. T. 11, pp. 105-127. Studies in Language Variation (cf. p. 36).

CORDEREIX, Pascal, 2001. Ferdinand Brunot, le phonographe et les « patois ». *Le Monde alpin et rhodanien: Revue régionale d'ethnologie*. T. 29, n° 1-3, pp. 39-54 (cf. p. 177, 178).

CORDEREIX, Pascal, 2014. Ferdinand Brunot et les Archives de la parole: le phonographe, la mort, la mémoire. *Revue de la BNF*. T. 48, n° 3, pp. 5-11 (cf. p. 186).

CÔTÉ, Marie-Hélène, 2005. Le statut lexical des consonnes de liaison. *Langages*. T. 158, n° 2, pp. 66-78 (cf. p. 118).

CÔTÉ, Marie-Hélène, 2011. French Liaison. In : *The Blackwell Companion to Phonology*. M. van Oostendorp, C. J. Ewen, E. Hume & K. Rice. Chichester, West Sussex : Wiley-Blackwell. T. V: Phonology Across Languages, pp. 1-26 (cf. p. 38).

CÔTÉ, Marie-Hélène, 2012. Laurentian French (Quebec): Extra vowels, missing schwas and surprising liaison consonants. In : *Phonological Variation in French: Illustrations from three continents*. R. Gess, Ch. Lyche & T. Meisenburg. Amsterdam; Philadelphia : John Benjamins. T. 11, pp. 235-274. Studies in Language Variation (cf. p. 118).

CÔTÉ, Marie-Hélène, 2017. La liaison en diatopie: esquisse d'une typologie. *Journal of French Language Studies*. Vol. 27, no. 1, pp. 13–25 (cit. on pp. 47, 51, 81, 82, 93, 117, 118, 240, 246).

COUTANSON, Gasparde, 2016. *Le pataquès: Nouvel éclairage via le prisme d'un corpus de chansons françaises*. Lyon. Mémoire de master 1: Sciences du langage. Université Lumière Lyon II (cf. p. 16, 104, 111, 115, 121, 179, 274).

COUTANSON, Gasparde, 2017. *Le pataquès: étude d'un cas de sandhi externe non normatif propre au français au sein d'un corpus de chansons*. Lyon. Mémoire de master 2: Sciences du langage. Université Lumière Lyon II (cf. p. 16, 17, 34, 40, 41, 102-106, 108, 110-113, 118, 121, 179, 180, 221, 269, 274).

COUTANSON, Gasparde, 2020. Mettre en place une méthodologie pour l'étude de phénomènes morphophonologiques rares: l'exemple du pataquès. *Linx: Revue des linguistes de l'Université Paris Ouest Nanterre La Défense*. N° 13 (cf. p. 121, 123, 125-129, 132, 133, 212, 239, 246).

COUTANSON, Gasparde & BADIN, Flora, 2021. La liaison dans un corpus de hits francophones (1956–2017). *Journal of French Language Studies*. T. 31, n° 2, pp. 131-147 (cf. p. 144, 148, 149, 151, 153, 184, 198-200, 263-265).

COUTANSON, Gasparde & BADIN, Flora, sous presse. Liaisons variables et pataquès dans un corpus de chansons de tradition orale. *Langue française* (cf. p. 79, 263, 265-267, 274, 276).

CRYSTAL, David, 1980. *A Dictionary of Linguistics and Phonetics*. 6ème édition (2008). Malden, MA ; Oxford ; Carlton, Victoria : Blackwell. The language library (cf. p. 19, 20).

DAWSON, Alain ; AURAN, Cyril ; BOUZON, Caroline ; DELRUE, Laurence ; LOOCK, Rudy ; O'CONNOR, Kathleen M. & PATIN, Cédric, 2016. French in Nord (Nord-Pas de Calais): A speaker from La Madeleine. In : *Varieties of Spoken French*. S. Detey, J. Durand, B. Laks & Ch. Lyche. Oxford : Oxford University Press, pp. 148-158. Oxford Linguistics (cf. p. 95).

DE JONG, Daan, 1994. La sociophonologie de la liaison orléanaise. In : *French Generative Phonology: Retrospective and Perspectives*. C. Lyche. Salford : Association for French Language Studies/European Studies Research Institute, pp. 95-130 (cf. p. 50).

DELATTRE, Pierre, 1947. La Liaison en Français, Tendances et Classification. *The French Review*. T. 21, n° 2, pp. 148-157 (cf. p. 30, 40, 41, 44, 46, 48, 49, 51, 54, 220, 222, 255, 257, 280, 296).

DELATTRE, Pierre, 1955. Les Facteurs de la Liaison Facultative en Français. *The French Review*. T. 29, n° 1, pp. 42-49 (cf. p. 45, 46, 48, 49, 85).

DELATTRE, Pierre, 1956. La Fréquence des liaisons facultatives en français. *The French Review*. T. 30, n° 1, pp. 48-54 (cf. p. 47, 49-51, 171).

DESAGULIER, Guillaume, 2017. *Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics*. Cham: Springer. Quantitative Methods in the Humanities and Social Sciences (cit. on p. 130).

DESGRUGILLERS, Éric, 2014. *Des chansons tissées aux fuseaux: L'art de Virginie Granouillet: Dentellière à Roche-en-Régnier*. Yssingeaux : Cahiers de la Haute-Loire/Agence des Musiques et Territoires d'Auvergne (cf. p. 182, 183, 192).

DESROCHERS, Richard, 1994. Les Liaisons Dangereuses: Le Statut Équivoque des Erreurs de Liaison. *Lingvisticae Investigationes*. T. 18, n° 2, pp. 243-284 (cf. p. 40, 65, 90, 100-102, 105-111, 116, 118, 121, 122).

DUGUA, Céline & BAUDE, Olivier, 2017. La liaison à Orléans, corpus et changement linguistique: une première étude exploratoire. *Journal of French Language Studies*. T. 27, n° 1, pp. 41-54 (cf. p. 69, 70, 95, 148).

DUGUA, Céline ; BAUDE, Olivier ; BADIN, Flora ; CÔTÉ, Marie-Hélène ; GANAYE, Jennifer ; LIÉGEOIS, Loïc ; PARISSE, Christophe & SICCARDI, Anne, 2017. La liaison dans l'environnement langagier des enfants: Vers une annotation commune? In : *Français Langue Orale et Recherches Avancées en Linguistique (FLORAL 2017): Accessibilité, représentations et analyses des données*. Orléans, France (cf. p. 32, 34, 70, 71, 134, 140).

DUGUA, Céline & GANAYE, Jennifer, 2022. *La question du lien input/output: usage des liaisons au sein de trois familles dans le corpus ESLO-Enfants*. Vienne, Autriche (cf. p. 140, 141).

DUNETON, Claude, 1998. *Histoire de la chanson française: Des origines à 1780*. T. 1. Paris, France : Seuil (cf. p. 178, 180).

DURAND, Jacques ; LAKS, Bernard ; CALDERONE, Basilio & TCHOBANOV, Atanas, 2011. Que savons-nous de la liaison aujourd'hui ? *Langue française*. T. 169, n° 1, pp. 103-135 (cf. p. 21, 27-29, 32, 34, 40, 46, 50, 82, 83, 88).

DURAND, Jacques & LYCHE, Chantal, 2003. Le projet 'Phonologie du français contemporain' (PFC) et sa méthodologie. In : *Corpus et variation en phonologie du français: Méthodes et analyses*. E. Delais-Roussarie & J. Durand. Toulouse : Presses Universitaires du Mirail. Interlangue: Linguistique et didactique (cf. p. 21).

DURAND, Jacques & LYCHE, Chantal, 2008. French liaison in the light of corpus data. *Journal of French Language Studies* [en ligne]. Vol. 18, no. 1, pp. 33–66 Consulté le 10/01/2023 (cit. on pp. 21, 32, 49, 50).

DURAND, Jacques & LYCHE, Chantal, 2016. Approaching variation in PFC: The liaison level. In : *Varieties of Spoken French*. S. Detey, J. Durand, B. Laks & Ch. Lyche. Oxford : Oxford University Press, pp. 363-375. Oxford Linguistics (cf. p. 16, 21).

ENCREVÉ, Pierre, 1983. La liaison sans enchaînement. *Actes de la recherche en sciences sociales*. T. 46, n° 1, pp. 39-66 (cf. p. 27, 33, 34, 38, 46, 47, 50-54, 65, 66, 70, 88, 90, 280).

ENCREVÉ, Pierre, 1988. *La liaison avec et sans enchaînement: Phonologie tridimensionnelle et usages du français*. Paris : Seuil. Travaux linguistiques (cf. p. 27, 66, 89).

ENGLEBERT, Annick, 2009. *Introduction à la phonétique historique du français*. Bruxelles : De Boeck Supérieur. Champs linguistiques (cf. p. 41, 55, 56, 60-64).

ESHKOL-TARAVELLA, Iris ; BAUDE, Olivier ; MAUREL, Denis ; HRIBA, Linda ; DUGUA, Céline & TELLIER, Isabelle, 2011. Un grand corpus oral «disponible»: le corpus d'Orléans 1968-2012. *Traitement Automatique des Langues*. T. 52, n° 3, pp. 17-46 (cf. p. 32, 69, 95, 203).

EYCHENNE, Julien & LAKS, Bernard, 2017. La liaison en français contemporain: normes, usages, acquisitions. *Journal of French Language Studies*. T. 27, n° 1, pp. 1-12 (cf. p. 22).

EYCHENNE, Julien ; NAVARRO, Sylvain ; TCHOBANOV, Atanas & VAN LEUSSEN, Jan-Willem, 2016. Approaching variation in PFC~: The tools. In : *Varieties of Spoken French*. S. Detey, J. Durand, B. Laks & Ch. Lyche. Oxford : Oxford University Press, pp. 389-400. Oxford Linguistics (cf. p. 210).

FORNEL, Michel de & VERDIER, Maud, 2018. Corpus, classes d'exemples et collections en analyse de conversation. *Corpus*. N° 18 (cf. p. 122).

FOUCHÉ, Pierre, 1956. *Traité de prononciation française*. 2ème édition (1959), réimpression (1988). Paris : Klincksieck (cf. p. 36, 37, 79, 254, 255, 257, 296).

FREI, Henri, 1929. *La grammaire des fautes: introduction à la linguistique fonctionelle, assimilation et différenciation, brièveté et invariabilité, expressivité*. Réimpression (1993). Genève; Paris : Slatkine Reprints (cf. p. 40, 90, 103, 108, 110, 111, 116, 122).

GAATONE, David, 1978. Forme sous-jacente unique ou liste d'allomorphes? (à propos des consonnes de liaison en français). *Linguistics*. T. 16, n° 214, pp. 33-54 (cf. p. 90).

GAATONE, David, 1979. Liaison et structure syllabique en français. *Le français moderne*. T. 47, n° 4, pp. 312-334 (cf. p. 90, 106).

GADET, Françoise, 1989. *Le français ordinaire*. 2ème édition (1997). Paris : Armand Colin. U (cf. p. 90, 98, 100, 101, 116, 169).

GADET, Françoise, 1992. *Le français populaire*. Paris : Presses Universitaires de France. Que sais-je?, n° 1172 (cf. p. 40, 90, 109, 116).

GADET, Françoise, 1995. Les relatives non standard en français parlé: le système et l'usage. In : *La subordination dans les langues romanes: Actes du colloque international Copenhague 5.5.-7.5. 1994*. Copenhague : Université de Copenhague. T. 34, pp. 141-162. Etudes Romanes (cf. p. 109).

GEORGETON, Laurianne ; PAILLEREAU, Nikola ; LANDRON, Simon ; GAO, Jiayin & KAMIYAMA, Takeki, 2012. Analyse formantique des voyelles orales du français en contexte isolé: à la recherche d'une référence pour les apprenants de FLE. In : *Actes de la conférence conjointe JEP-TALN-RECITAL 2012*. Grenoble : Association Francophone pour la Communication Parlée (AFCP); Association pour le Traitement Automatique des Langues (ATALA). T. 1: JEP, pp. 145-152 (cf. p. 251).

GESS, Randall ; LYCHE, Chantal & MEISENBURG, Trudel, 2012. Introduction to phonological variation in French: Illustrations from three continents. In : *Phonological Variation in French: Illustrations from three continents*. R. Gess, Ch. Lyche & T. Meisenburg. Amsterdam; Philadelphia : John Benjamins. T. 11, pp. 1-19. Studies in Language Variation (cf. p. 57).

GOLDBERG, Adele E., 2003. Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*. T. 7, n° 5, pp. 219-224 (cf. p. 99, 252).

GRAMMONT, Maurice, 1933. *Traité de phonétique: avec 179 figures dans le texte*. 3ème édition (1946). Paris : Delagrave (cf. p. 22, 65).

GRAVIER, Guillaume; ADDA, Gilles; PAULSON, Niklas; CARRÉ, Matthieu; GIRAUDEL, Aude & GALIBERT, Olivier, 2012. The ETAPE corpus for the evaluation of speech-based TV content processing in the French language. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation*. Istanbul, Turquie, pp. 114–118 (cit. on p. 67).

GREEFHORST, Alexander & VAN DEN BOSCH, Antal, 2016. Predicting Liaison: an Example-Based Approach. *Traitement Automatique des Langues*. Vol. 57, no. 1, pp. 13–32 (cit. on p. 145).

HABERT, Benoît, 2000. Des corpus représentatifs: de quoi, pour quoi, comment? *Cahiers de l'Université de Perpignan*. N° 31, pp. 11-58 (cf. p. 132, 285).

HABERT, Benoît; NAZARENKO, Adeline & SALEM, André, 1997. *Les linguistiques de corpus*. Paris : Armand Colin. U (cf. p. 127, 130).

HANSEN, Anita Berit, 2016. French in Paris (Île-de-France): A speaker from the 14th arrondissement. In : *Varieties of Spoken French*. S. Detey, J. Durand, B. Laks & Ch. Lyche. Oxford : Oxford University Press, pp. 123-136. Oxford Linguistics (cf. p. 81).

HORNSBY, David, 2012. Getting it Wrong: Liaison, Pataquès, and Repair in Contemporary French. In: *On linguistic change in French: socio-historical approaches: Studies in honour of R. Anthony Lodge/Le changement linguistique en français: aspects socio-historiques: Études en hommage au Professeur R. Anthony Lodge*. T. Pooley & D. Lagorgette. Chambéry: Université de Savoie, pp. 69–84. Langages (cit. on pp. 87, 88, 90, 103, 120, 121, 282, 283).

HORNSBY, David, 2019. Variable liaison, diglossia, and the style dimension in spoken French. *French Studies*. Vol. LXXIII, no. 4, pp. 578–597 (cit. on pp. 83, 87).

HORNSBY, David, 2020. *Norm and Ideology in Spoken French: A Sociolinguistic History of Liaison*. Cham: Palgrave Macmillan (cit. on pp. 62, 73, 79, 83, 87, 103, 119, 120, 250, 284).

HUTIN, Mathilde; WENG, Caihong; ADDA-DECKER, Martine & LAMEL, Lori, 2022. La liaison facultative en français: étude de grands corpus combinant approche automatique relâchée et jugement perceptif. In : *8e Congrès Mondial de Linguistique Française 2022*. Orléans, France : EDP Sciences. T. 138 (cf. p. 67, 71, 90, 149, 150, 287).

INSTITUT PIERRE GARDETTE (IPG) & DYNAMIQUE DU LANGAGE - UMR 5596 (DDL), 2017. *Atlas Linguistique et ethnographique du Lyonnais (ALLy) [Corpus]* [en ligne]. Consulté le 15/03/2023. Disponible à l'adresse : https://hdl.handle.net/11403/ally.

JONES, Mari C., 2012. Liaison Patterns and Usage in Jersey Norman French. *Probus*. T. 24, nº 2, pp. 197-232 (cf. p. 95, 125).

KOCH, Peter & OESTERREICHER, Wulf, 2001. Langage parlé et langage écrit. In : *Lexikon der Romanistischen Linguistik (LRL)*. G. Holtus, M. Metzeltin & Ch. Schmitt. Tübingen : Max Niemeyer Verlag. T. I, 2: Methodologie (Sprache in der Gesellschaft / Sprache und Klassifikation / Datensammlung und -verarbeitung)/Méthodologie (Langue et société / Langue et classification / Collection et traitement des données), pp. 584-627 (cf. p. 51, 197, 245, 262, 268).

KRISTOL, Andres Max, 1998. La production interactive d'un corpus semi-spontané: l'expérience ALAVAL. *Cahiers du Centre de Linguistique et des Sciences du Langage*. Nº 10, pp. 91-104 (cf. p. 203).

KURSELL, Julia ; TKACZYK, Viktoria & ZIEMER, Hansjakob, 2021. Introduction: Language, Sound, and the Humanities. *History of Humanities*. T. 6, nº 1, pp. 1-10 (cf. p. 177).

LABORDERIE, Noëlle, 1994. *Précis de phonétique historique*. 2ème édition (2009). Paris : Armand Colin. 128 (cf. p. 57, 60-64, 79, 109).

LABOV, William, 1976. *Sociolinguistique*. Trad. par KIHM, Alain. Paris : Éditions de Minuit. Le sens commun (cf. p. 124).

LABOV, William, 1994. *Principles of Linguistic Change: Internal Factors*. Vol. 1. Réimpression (1999). Malden, MA; Oxford: Blackwell. Language in Society (cit. on p. 66).

LAFORTE, Conrad, 1976. *Poétiques de la chanson traditionnelle française: Classification de la chanson folklorique française*. 2ème édition (1993). Sainte-Foy : Les Presses de l'Université Laval. Les archives de folklore, nº 26 (cf. p. 178, 179).

LAKS, Bernard, 2005a. La liaison et l'illusion. *Langages*. Nº 2, pp. 101-125 (cf. p. 20, 41, 53).

LAKS, Bernard, 2005b. Phonologie et construction syntaxique: la liaison, un test de cohésion et de figement syntaxique. *Linx: Revue des linguistes de l'université Paris X Nanterre*. Nº 53, pp. 155-171 (cf. p. 45, 47, 48).

LAKS, Bernard, 2008. Dynamiques de la liaison en français. In : *Le français d'un continent à l'autre: Mélanges offerts à Yves Charles Morin*. L. Baronian & F. Martineau. Québec : Les Presses de l'Université Laval, pp. 237-267. Les voies du français (cf. p. 27, 47, 48, 50, 52, 66-71, 142).

LAKS, Bernard, 2014. Diachronie de la liaison en français contemporain: le cas de la parole publique (1999-2011). In : *La phonologie du français: normes, périphéries, modélisation: Mélanges pour Chantal Lyche*. J. Durand, G. Kristoffersen & B. Laks. Nanterre : Presses Universitaires de Paris Ouest, pp. 327-379. Les voies du français (cf. p. 53, 66-68).

LAKS, Bernard & PEUVERGNE, Julie, 2017. La liaison en français contemporain dans la parole publique (1999–2015). *Journal of French Language Studies*. T. 27, nº 1, pp. 55-72 (cf. p. 66-71, 203).

LEECH, Geoffrey, 1991. The state of the art in corpus linguistics. In: *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. K. Aijmer & B. Altenberg. New York; London: Routledge [2013], pp. 8–29 (cit. on pp. 124, 151).

LEHMANN, Christian, 2004. Interlinear morphemic glossing. In : *Morphologie: Ein internationales Handbuch zur Flexion und Wortbildung/Morphology: An International Handbook on Inflection and Word-Formation*. G. E. Booij, Ch. Lehmann, J. Mugdan & S. Skopeteas. Berlin; New York : Walter de Gruyter. T. 2, pp. 1834-1857. Handbücher der Sprach- und Kommunikationswissenschaft/ Handbooks of Linguistics and Communication Science (cf. p. 335).

LÉON, Monique, 1984. Erreurs et normalisation: les liaisons fautives en français contemporain. *Revue de Phonétique Appliquée*. Nº 69, pp. 1-9 (cf. p. 40, 44, 90, 104-106, 116, 121).

LÉON, Pierre R., 1992. *Phonétisme et prononciations du français: avec travaux pratiques d'application et corrigés*. 6ème édition (2011). Paris : Armand Colin. Cursus Lettres (cf. p. 41, 42, 54, 65, 79, 90, 116).

LÉON, Pierre R., 1993. *Précis de phonostylistique: parole et expressivité*. Paris : Nathan. Fac: Linguistique (cf. p. 116).

LIÉGEOIS, Loïc, 2014. *Usage des variables phonologiques dans un corpus d'interactions naturelles parents-enfant: impact du bain linguistique et dispositifs cognitifs d'apprentissage.* Clermont-Ferrand. Thèse de doctorat: Sciences du langage. Université Blaise Pascal – Clermont Université (cf. p. 100, 134, 141, 142, 157, 207, 208, 226, 228, 229, 231, 233, 237, 240, 311, 315).

LIÉGEOIS, Loïc ; ETIENNE, Carole ; PARISSE, Christophe ; BENZITOUN, Christophe & CHANARD, Christian, 2015. Using the TEI as a pivot format for oral and multimodal language corpora. In : *Text Encoding Initiative Conference and Member's meeting 2015*. Lyon, France, pp. 28-31 (cf. p. 144, 223).

LODGE, Anthony, 2011. La question de la «langue commune» en français: Normes «sociales» vs normes «communautaires». In : *Langue commune et changements de normes*. S. Branca-Rosoff, J.-M. Fournier, Y. Grinshpun & A. Régent-Susini. Paris : Honoré Champion, pp. 77-92. Linguistique historique, n° 2 (cf. p. 91).

MACWHINNEY, Brian, 2000a. *The CHILDES Project: Tools for Analyzing Talk*. T. I: Transcription Format and Programs. 3ème édition (2014). New York : Psychology Press (cf. p. 134, 135, 137, 138, 141, 223).

MACWHINNEY, Brian, 2000b. *The CHILDES Project: Tools for Analyzing Talk*. T. II: the Database. 3ème édition (2014). New York : Psychology Press (cf. p. 135).

MALÉCOT, André, 1975. French Liaison as a Function of Grammatical, Phonetic and Paralinguistic Variables. *Phonetica*. T. 32, n° 3, pp. 161-179 (cf. p. 30, 32-35, 50, 268).

MALLET, Géraldine-M., 2008. *La liaison en français: descriptions et analyses dans le corpus PFC*. Nanterre. Thèse de doctorat: Sciences du langage. Université Paris Ouest Nanterre La Défense (cf. p. 23, 24, 33, 35, 45, 46, 48-51, 98, 99, 268, 279).

MARCADÉ, Claire ; GUINARD, Bernard ; COULAIS, Stéphanie ; DAVY, Yvon ; DESGRUGILLERS, Eric ; GASNAULT, François ; GINOUVÈS, Véronique ; MARCOTTE, Pierre ; O'SULLIVAN, Mikaël ; RAMEL, Jean-Luc ; SARRAT, Gwenaëlle & THOMAS, Laura, 2014. *Patrimoine culturel immatériel: Traitement documentaire des archives sonores inédites: Guide des bonnes pratiques*. Fédération des Associations de Musiques et Danses Traditionnelles (cf. p. 159, 180).

MARCHELLO-NIZIA, Christiane ; COMBETTES, Bernard ; PRÉVOST, Sophie & SCHEER, Tobias (éd.), 2020. *Grande grammaire historique du français*. T. 1. Berlin; Boston : Walter de Gruyter GmbH (cf. p. 41, 43, 56-58, 62, 63, 71, 73-76, 80, 90, 95, 169, 170).

MARTINET, André, 1968. Neutralisation et syncrétisme. *La linguistique*. T. 4, n° 1, pp. 1-20 (cf. p. 42).

MARTINON, Philippe, 1913. *Comment on prononce le français: Traité complet de prononciation pratique avec les noms propres et les mots étrangers*. Paris : Larousse (cf. p. 82, 90).

MAZALEYRAT, Jean, 1974. *Éléments de métrique française*. 8ème édition (2016). Malakoff : Armand Colin. Cursus (cf. p. 84, 85).

MCCARTHY, Michael & O'KEEFFE, Anne, 2010. Historical perspective: What are corpora and how have they evolved? In: *The Routledge Handbook of Corpus Linguistics*. A. O'Keeffe & M. McCarthy. London; New York: Routledge, pp. 3–13. Routledge Handbooks in Applied Linguistics (cit. on p. 124).

MICHEL, Jean-Baptiste ; SHEN, Yuan Kui ; AIDEN, Aviva Presser ; VERES, Adrian ; GRAY, Matthew K. ; TEAM, The Google Books ; PICKETT, Joseph P. ; HOIBERG, Dale ; CLANCY, Dan ; NORVIG, Peter ; ORWANT, Jon ; PINKER, Steven ; NOWAK, Martin A. & AIDEN, Erez Lieberman, 2011. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*. T. 331, nº 6014, pp. 176-182 (cf. p. 23).

MILNER, Jean-Claude & REGNAULT, François, 1987. *Dire le vers: court traité à l'intention des acteurs et des amateurs d'alexandrins*. Paris : Seuil (cf. p. 27, 84).

MILROY, James, 2006. The ideology of the standard language. In : *The Routledge Companion to Sociolinguistics*. C. Llamas, L. Mullany & P. Stockwell. London; New York : Routledge, pp. 133-139 (cf. p. 93).

MOMPEÁN-GONZALEZ, Jose & MOMPEÁN-GUILLAMÓN, Pilar, 2009. /r/-liaison in English: An empirical study. *Cognitive Linguistics*. Vol. 20, no. 4, pp. 733–776 (cit. on p. 21).

MORIN, Yves-Charles, 1982. De quelques [l] non étymologiques dans le français du Québec: notes sur les clitiques et la liaison. *Revue québécoise de linguistique*. T. 11, nº 2, pp. 9-47 (cf. p. 29, 59, 94, 102, 118, 224).

MORIN, Yves-Charles, 1986. On the morphologization of word-final consonant deletion in French. In : *Sandhi phenomena in the languages of Europe*. H. Andersen. Berlin; New York; Amsterdam : Mouton de Gruyter, pp. 167-210 (cf. p. 38, 42, 43, 61, 71, 72, 108).

MORIN, Yves Charles, 1987. Remarques sur l'organisation de la flexion des verbes français. *ITL Review of Applied Linguistics*. T. 77, nº 78, pp. 13-91 (cf. p. 29).

MORIN, Yves-Charles, 1992. Un cas méconnu de la déclinaison de l'adjectif français: les formes de liaison de l'adjectif antéposé. In : *Le mot, les mots, les bons mots/Word, words, witty words*. A. Clas. Montréal : Les Presses de l'Université de Montréal, pp. 233-250 (cf. p. 39, 40).

MORIN, Yves Charles, 2005. La liaison relève-t-elle d'une tendance à éviter les hiatus? Réflexions sur son évolution historique. *Langages*. T. 2, n° 158, pp. 8-23 (cf. p. 24, 25, 58, 125, 278).

MORIN, Yves Charles, 2019. L'augment clitique le dans le français du Poitou à la fin du XVIe et au début du XVIIe siècle : un effet du substrat ? *Langages* [en ligne]. T. 215, n° 3, pp. 73-88 Consulté le 05/04/2023 (cf. p. 110).

MORIN, Yves-Charles & BONIN, Michèle, 1992. Les -S analogiques des 1sg au XVIe siècle: les témoignages de Meigret et Lanoue. *Revue québécoise de linguistique*. T. 21, n° 2, pp. 33-63 (cf. p. 74).

MORIN, Yves Charles & KAYE, Jonathan D., 1982. The syntactic bases for French liaison. *Journal of Linguistics*. Vol. 18, no. 2, pp. 291–330 (cit. on pp. 40, 45, 47, 50, 106, 108, 109, 111, 123, 251, 274).

MÜLLER, Bodo, 1985. *Le français d'aujourd'hui*. Trad. par ELSASS, Annie. Paris : Klincksieck. Bibliothèque française et romane: Manuels et études linguistiques, n° 47 (cf. p. 22, 30, 65, 90, 116).

NARDY, Aurélie ; CHEVROT, Jean-Pierre & CHAUVIN, Carole, 2014. La liaison facultative dans les formes récitées du folklore enfantin. In : *La liaison: approches contemporaines*. Ch. Soum-Favaro, A. Coquillon & J.-P. Chevrot. Bern; Berlin; Bruxelles; Frankfurt am Main; New York; Oxford; Wien : Peter Lang. T. 110, pp. 239-262. Sciences pour la communication (cf. p. 86, 90, 116, 149, 262-264, 282).

NEW, Boris ; PALLIER, Christophe ; FERRAND, Ludovic & MATOS, Rafael, 2001. Une base de données lexicales du français contemporain sur internet: LEXIQUE™//A lexical database for contemporary french: LEXIQUE™. *L'année psychologique*. T. 101, n° 3, pp. 447-462 (cf. p. 145).

PERNOT, Nicolette, 1937. Quelques Notes sur la liaison en français: Liaison et Enchaînement. *The Modern Language Journal*. T. 21, n° 5, pp. 333-338 (cf. p. 26).

PERRET, Michèle, 1998. *Introduction à l'histoire de la langue française*. 3ème édition (2008). Paris : Armand Colin. Cursus (cf. p. 57, 58, 73, 79).

PICHON, Édouard, 1935. L'enrichissement lexical dans le français d'aujourd'hui (suite et fin). *Le français moderne*. T. 3, n° 4, pp. 325-344 (cf. p. 101, 108).

PLÉNAT, Marc, 2008. La liaison «obligatoire» avec et sans enchaînement. In : *Congrès Mondial de Linguistique Française 2008*. Paris, France : EDP Sciences (cf. p. 39, 40).

PLÉNAT, Marc & PLÉNAT, Camille, 2011. La liaison de l'adjectif sur le nom en français: morphologie, syntaxe, phonologie. *Linguistica*. T. 51, n° 1, pp. 299-315 (cf. p. 43).

PUSTKA, Elissa, 2009. PFC et la phonologie du français en Vendée. In : *Phonologie, variation et accents du français*. J. Durand, B. Laks & Ch. Lyche. Paris : Hermès, pp. 307-335. Traité IC2: Cognition et traitement de l'information (cf. p. 95).

PUSTKA, Elissa, 2017. L'écrit avant l'écriture: la liaison dans les livres audio pour enfants. *Journal of French Language Studies* [en ligne]. T. 27, n° 2, pp. 187-214 Consulté le 25/10/2018 (cf. p. 47, 49, 50, 83).

RACINE, Isabelle & ANDREASSEN, Helene N, 2012. A phonological study of a Swiss French variety: Data from the canton of Neuchâtel. In : *Phonological Variation in French: Illustrations from three continents*. R. Gess, Ch. Lyche & T. Meisenburg. Amsterdam; Philadelphia : John Benjamins. T. 11, pp. 173-207. Studies in Language Variation (cf. p. 34, 82).

REECE, Steve, 2009. Junctural Metanalysis In Middle English. In : *Homer's Winged Words: The Evolution of Early Greek Epic Diction in the Light of Oral Theory*. Leiden; Boston : Brill, pp. 15-26. Mnemosyne supplements: monographs on Greek and Roman language and literature, n° 313 (cf. p. 64).

ROCHÉ, Michel, 2010. Base, thème, radical. *Recherches linguistiques de Vincennes*. N° 39, pp. 95-134 (cf. p. 38, 39).

ROCHER, Claude, 1994. *En écoutant la baracande: Avant le grand silence...* Aubenas : Fédération Départementale des Foyers Ruraux de l'Allier (cf. p. 183, 185-190, 192, 193, 195).

RUSSO, Michela, 2014. *Les origines de la liaison en proto-français: graphies et oralité*. Cambridge, Royaume-Uni (cf. p. 57).

RUSSO, Michela & SAHMAOUI, Laure, 2021. Tu l'as vu? [tyllavy]: La gémination des proclitiques dans les langues de France. *Linx*. T. 82 (cf. p. 109).

SAUSSURE, Ferdinand (de), 1916. *Cours de linguistique générale (1916)*. Sous la dir. de BALLY, Charles & SECHEHAYE, Albert. Paris : Payot (cf. p. 251).

STEIN, Achim & SCHMID, Helmut, 1995. Etiquetage morphologique de textes français avec un arbre de décisions. *Traitement automatique des langues*. T. 36, n° 1-2, pp. 23-35 (cf. p. 145, 335).

STICH, Dominique, 1998. *Parlons francoprovençal: Une langue méconnue*. Paris; Montréal : L'Harmattan. Parlons (cf. p. 125).

SURMONT, Jean-Nicolas de, 2010. *Vers une théorie des objets-chansons*. Lyon : ENS Éditions. Signes (cf. p. 178, 179, 196, 197).

THOGMARTIN, Clyde O., 1970. *The French dialect of Old Mines, Missouri*. Ann Arbor. Thèse de doctorat. University of Michigan (cit. on p. 95).

THOMAS, Alain, 1998. La liaison et son enseignement: des modèles orthoépiques à la réalité linquistique. *The Canadian Modern Language Review/La revue canadienne des langues vivantes*. T. 54, nº 4, pp. 543-552 (cf. p. 65).

TORREIRA, Francisco ; ADDA-DECKER, Martine & ERNESTUS, Mirjam, 2010. The Nijmegen Corpus of Casual French. *Speech Communication*. T. 52, nº 3, pp. 201-212 (cf. p. 32).

TRANEL, Bernard, 1987. Liaison. In : *The sounds of French: An introduction*. Cambridge ; New York ; Melbourne : Cambridge University Press, pp. 168-190 (cf. p. 40, 65, 90, 274).

TRANEL, Bernard, 1990. On Suppletion and French Liaison. *Probus*. Vol. 2, no. 2, pp. 169–208 (cit. on p. 38).

TRANEL, Bernard, 2000. Aspects de la phonologie du français et la théorie de l'optimalité. *Langue française*. Nº 126, pp. 39-72 (cf. p. 24).

WALKER, Douglas C., 2012. Albertan French phonology: French in an anglophone context. In : *Phonological Variation in French: Illustrations from three continents*. R. Gess, Ch. Lyche & T. Meisenburg. Amsterdam; Philadelphia : John Benjamins. T. 11, pp. 341-368. Studies in Language Variation (cf. p. 82).

WALTER, Henriette, 1988. *Le français dans tous les sens: Grandes et petites histoires de notre langue*. (2008). Paris : Points. Le goût des mots (cf. p. 165).

WAUQUIER, Sophie, 2009. Acquisition de la liaison en L1 et L2: stratégies phonologiques ou lexicales? *Acquisition et interaction en langue étrangère*. Nº Aile... Lia 2, pp. 93-130 (cf. p. 46).

WILL, Udo, 1999. La baguette magique de l'ethnomusicologue: Repenser la notation et l'analyse de la musique. *Cahiers d'ethnomusicologie (anciennement Cahiers de musiques traditionnelles)*. Nº 12, pp. 9-32 (cf. p. 189, 190).

ZEMP, Hugo, 1996. The/An Ethnomusicologist and the Record Business. *Yearbook for Traditional Music*. T. 28, pp. 36-56 (cf. p. 181).

ZUMTHOR, Paul, 1986. Chansons «médiatisées». *Études françaises*. T. 22, nº 3, pp. 13-19 (cf. p. 197).

Cours universitaires

HAMBYE, Philippe, Année universitaire 2016-2017. *Linguistique française II: Approche historique (ROM1222), Faculté de philosophie, arts et lettres, Université catholique de Louvain* (cf. p. 60, 62).

THIBAULT, André, Année universitaire 2019-2020. *Phonologie: semaine 12 (M1FA76LF), Master 1 Langue Française Appliquée, Faculté des Lettres, Sorbonne Université* (cf. p. 63).

THIBAULT, André, Année universitaire 2021-2022. *Phonologie: module 14 (M2FA76LF), Master 1 Langue Française Appliquée, Faculté des Lettres, Sorbonne Université* (cf. p. 57).

TUTIN, Agnès ; JACQUES, Marie-Paule ; KRAIF, Olivier & HARTWELL, Laura, 2019. *Les principaux types de corpus. Introduction à la linguistique de corpus, MOOC, plateforme Fun MOOC, Université Grenoble Alpes*. Disponible à l'adresse : https://www.fun-mooc.fr/fr/cours/introduction-a-la-linguistique-de-corpus/ (cf. p. 127).

Corpus

BORÉ, Catherine ; ROUBAUD, Marie-Noëlle & ELALOUF, Marie-Laure, 2017. *Corpus ÉMA, écrits scolaires [Corpus]*. Disponible à l'adresse : https://hdl.handle.net/11403/ema-ecrits-scolaires-1 (cf. p. 77).

CHABANAL, Damien ; CHANIER, Thierry & LIÉGEOIS, Loïc, 2017. *ALIPE (Acquisition de la Liaison et Interactions Parents Enfants) [Corpus]*. Disponible à l'adresse : https://hdl.handle.net/11403/alipe-000853/v1.1 (cf. p. 207).

MODÈLES, DYNAMIQUES, CORPUS - UMR 7114 (MODYCO) & UNIVERSITÉ DE GRONINGEN (RUG), 2017. *PFC - Phonologie du Français Contemporain [Corpus]*. Disponible à l'adresse : https://hdl.handle.net/11403/pfc/v1 (cf. p. 206).

PRAXILING - UMR 5267, 2019. *Corpus 14 [Corpus]*. Disponible à l'adresse : https://hdl.handle.net/11403/corpus14/v2 (cf. p. 77).

Programmes, logiciels et outils informatiques, manuels

BADIN, Flora, 2019. *Liaisons sous Python [Outil]* (cf. p. 144).

BOERSMA, Paul & WEENINK, David, 1992. *Praat: doing phonetics by computer [Programme informatique]*. [en ligne]. Consulté le 08/09/2022. Disponible à l'adresse : https://www.praat.org (cf. p. 134, 223).

BOURDOUX, Françoise ; ROSSI, Caroline ; MORGENSTERN, Aliyah ; PARISSE, Christophe & CAËT, Stéphanie, 2011. *Guide d'utilisation de CLAN - Projet COLAJE* (cf. p. 134).

DURAND, Jacques ; LAKS, Bernard & LYCHE, Chantal, 2002. *Bulletin PFC n°1: Protocole d'enquête*. Nicole Serna (cf. p. 83, 142, 143, 207, 241).

LABORATOIRE LORRAIN DE RECHERCHE EN INFORMATIQUE ET SES APPLICATIONS - UMR 7503 (LORIA), 2016. *Jtrans [Outil]*. Disponible à l'adresse : https://hdl.handle.net/11403/jtrans/v1 (cf. p. 144).

MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS, 2022. *ELAN [Programme informatique]*. Nijmegen. Disponible à l'adresse : https://archive.mpi.nl/tla/elan (cf. p. 144).

MODÈLES, DYNAMIQUES, CORPUS - UMR 7114 (MODYCO), 2016. *teicorpo [Outil]*. Disponible à l'adresse : https://hdl.handle.net/11403/teicorpo/v1 (cf. p. 144, 223).

NEW, Boris & PALLIER, Christophe, [s. d.]. *Manuel de Lexique 3. Document version 3.11*. Disponible à l'adresse : lexique.org/_documentation/Manuel_Lexique.3.2.pdf (cf. p. 145).

SCHMID, Helmut, 2022. *TreeTagger: a part-of-speech tagger for many languages [Programme informatique]*. Munich. Disponible à l'adresse : https://cis.uni-muenchen.de/~schmid/tools/TreeTagger/ (cf. p. 144).

STEIN, Achim, 2003. *French TreeTagger Part-of-Speech Tags [Script]*. Disponible à l'adresse : https://cis.uni-muenchen.de/~schmid/tools/TreeTagger/ (cf. p. 107, 145, 263, 264, 275, 335).

SYNDICAT NATIONAL DE L'ÉDITION PHONOGRAPHIQUE, 2021. *Règles applicables aux tops*. Disponible à l'adresse : https://snepmusique.com/top-radios/a-propos-des-tops/ (cf. p. 153).

uMap [Logiciel], 2023 [en ligne]. Consulté le 25/06/2023. Disponible à l'adresse : https://umap.openstreetmap.fr/fr/ (cf. p. 242).

VIRY, Matthieu ; GIRAUD, Timothée & LAMBERT, Nicolas, 2023. *Magrit [Logiciel]* [en ligne]. Consulté le 22/06/2023. Disponible à l'adresse : http://magrit.cnrs.fr/ (cf. p. 242).

Dictionnaires, entrées de dictionnaires

CNRTL, 2012a. *AVOIR : Etymologie de AVOIR* [en ligne]. Consulté le 29/12/2022. Disponible à l'adresse : https://www.cnrtl.fr/etymologie/avoir (cf. p. 63).

CNRTL, 2012b. *EPITHESE: Définition de EPITHESE* [en ligne]. Consulté le 30/06/2021. Disponible à l'adresse : https://www.cnrtl.fr/definition/epithese (cf. p. 94).

CNRTL, 2012c. *FRANC ÉTRIER: Définition de FRANC ÉTRIER* [en ligne]. Consulté le 26/07/2021. Disponible à l'adresse : https://www.cnrtl.fr/definition/franc%20%C3%A9trier (cf. p. 29).

CNRTL, 2012d. *GRIOTTE: Etymologie de GRIOTTE* [en ligne]. Consulté le 15/07/2022. Disponible à l'adresse : https://www.cnrtl.fr/etymologie/griotte (cf. p. 64).

CNRTL, 2012e. *LIERRE: Etymologie de LIERRE* [en ligne]. Consulté le 15/07/2022. Disponible à l'adresse : https://www.cnrtl.fr/etymologie/lierre (cf. p. 64).

CNRTL, 2012f. *LUETTE: Etymologie de LUETTE* [en ligne]. Consulté le 15/07/2022. Disponible à l'adresse : https://www.cnrtl.fr/etymologie/luette (cf. p. 64).

CNRTL, 2012g. *NOMBRIL: Etymologie de NOMBRIL* [en ligne]. Consulté le 15/07/2022. Disponible à l'adresse : https://www.cnrtl.fr/etymologie/nombril (cf. p. 64).

CNRTL, 2012h. *ORANGE: Etymologie de ORANGE* [en ligne]. Consulté le 15/07/2022. Disponible à l'adresse : https://www.cnrtl.fr/etymologie/orange (cf. p. 64).

CNRTL, 2012i. *PARAGOGIQUE: Définition de PARAGOGIQUE* [en ligne]. Consulté le 30/06/2021. Disponible à l'adresse : https://www.cnrtl.fr/definition/paragogique (cf. p. 94).

CNRTL, 2012j. *PERDRIX: Etymologie de PERDRIX* [en ligne]. Consulté le 15/07/2022. Disponible à l'adresse : https://www.cnrtl.fr/etymologie/perdrix (cf. p. 64).

CNRTL, 2012k. *PIED-À-TERRE: Etymologie de PIED-À-TERRE* [en ligne]. Consulté le 27/12/2022. Disponible à l'adresse : https://www.cnrtl.fr/etymologie/pied-%C3%A0-terre (cf. p. 58).

CNRTL, 2012l. *PSILOSE : Définition de PSILOSE* [en ligne]. Consulté le 14/06/2023. Disponible à l'adresse : https://www.cnrtl.fr/definition/psilose (cf. p. 63).

CNRTL, 2012m. *TANTE: Etymologie de TANTE* [en ligne]. Consulté le 15/07/2022. Disponible à l'adresse : https://www.cnrtl.fr/etymologie/tante (cf. p. 64).

CNRTL, 2012n. *ZIEUTER : Etymologie de ZIEUTER* [en ligne]. Consulté le 16/02/2023. Disponible à l'adresse : https://www.cnrtl.fr/etymologie/zieuter (cf. p. 53).

HUET, Gérard, 2014. □□□□ *saṃdhi* [en ligne]. Version 3.05. Consulté le 09/04/2018. Disponible à l'adresse : http://sanskrit.inria.fr/DICO/67.html#sandhi (cf. p. 19).

PAULEAU, Christine, 2022. *1424 ZOREILLE* [en ligne]. Paris : Site du laboratoire Modèles Dynamiques Corpus (MoDyCo), UMR 7114 du CNRS/Université Paris Nanterre Consulté le 15/02/2023. Disponible à l'adresse : https://ressources.modyco.fr/dicocaledonien/ (cf. p. 54).

TRASK, Robert L., 1996. *A Dictionary of Phonetics and Phonology*. Édition digitale (2006). London; New York: Routledge (cit. on p. 80).

VALDMAN, Albert; ROTTET, Kevin J.; ANCELET, Barry J.; GUIDRY, Richard; KLINGLER, Thomas A.; LAFLEUR, Amanda; LINDNER, Tamara; PICONE, Michael D. & RYON, Dominique, 2010. *Dictionary of Louisiana French: as spoken in Cajun, Creole, and American Indian communities*. Jackson : University Press of Mississippi (cf. p. 118).

Sitographie

BIPO, 2022a. *Base inter-régionale du patrimoine oral* [en ligne]. Consulté le 26/07/2022. Disponible à l'adresse : http://patrimoine-oral.org/dyn/portal/index.seam?page=home (cf. p. 180).

BIPO, 2022b. *Chansons de couteliers : les enquêtes de Jean Dumas dans les Bois Noirs et la vallée de la Durolle (fonds Jean Dumas)* [en ligne]. Consulté le 24/07/2022. Disponible à l'adresse : http://patrimoine-oral.org/dyn/portal/index.seam?page=alo&aloId=12196&cid=3894 (cf. p. 183, 205).

BIPO, 2022c. *Chansons de dentellières: les enquêtes de Jean Dumas auprès des dentellières de la vallée de la Loire (fonds Jean Dumas)* [en ligne]. Consulté le 26/07/2022. Disponible à l'adresse : http://patrimoine-oral.org/dyn/portal/index.seam?page=alo&aloId=32165&cid=652 (cf. p. 183, 205).

BIPO, 2022d. *Chansons et danses de la Montagne Bourbonnaise : les enquêtes de Jean Dumas dans l'Allier (03)* [en ligne]. Consulté le 26/07/2022. Disponible à l'adresse : http://patrimoine-oral.org/dyn/portal/index.seam?page=alo&aloId=41402&cid=4175 (cf. p. 183).

BIPO, 2022e. *Fonds Charles Joisten* [en ligne]. Consulté le 28/11/2022. Disponible à l'adresse : http://patrimoine-oral.org/dyn/portal/index.seam?aloId=31770&page=alo&fonds=&cid=1070 (cf. p. 209).

BIPO, 2022f. *Fonds Jean Dumas* [en ligne]. Consulté le 29/07/2022. Disponible à l'adresse : http://patrimoine-oral.org/dyn/portal/index.seam;jsessionid=a61922a9f08c28d170aa14ef9
page=alo&aloId=12151&fonds=&cid=2578 (cf. p. 183).

BIPO, 2022g. *Les chansons de la Baracande : les enquêtes de Jean Dumas auprès de Virginie Granouillet (fonds Jean Dumas)* [en ligne]. Consulté le 24/07/2022. Disponible à l'adresse : http://patrimoine-oral.org/dyn/portal/index.seam?page=alo&aloId=12178&cid=3894 (cf. p. 160, 183, 205).

BIPO, 2022h. *Voix de l'Emblavez et du Velay: les enquêtes de Jean Dumas dans la Haute-Loire (fonds Jean Dumas)* [en ligne]. Consulté le 26/07/2022. Disponible à l'adresse : http://patrimoine-oral.org/dyn/portal/index.seam?page=alo&aloId=12197&cid=3894 (cf. p. 183, 205).

CMTRA, 2017. *Les chants de mai du Mont Pilat* [en ligne]. Consulté le 11/03/2021. Disponible à l'adresse : https://cmtra.hypotheses.org/1844 (cf. p. 188).

CONNECTED PAPERS, 2022. *Connected Papers: Les Liaisons Dangereuses: Le Statut Équivoque des Erreurs de Liaison* [en ligne]. Consulté le 17/08/2022. Disponible à l'adresse: https://www.connectedpapers.com/main/92bbbd3d0e1cd04fcebbcb6232b379f75519b700
Les-Liaisons-Dangereuses%3A-Le-Statut-%C3%89quivoque-des-Erreurs-de-Liaison/graph.

GOUVERNEMENT, 2021. *Réformer l'organisation territoriale: La réforme territoriale* [en ligne]. Consulté le 08/03/2021. Disponible à l'adresse : https://www.gouvernement.fr/action/la-reforme-territoriale (cf. p. 180).

LESUEUR, Daniel & DURAND, Dominic, 2022a. *Chronologie des différents Classements de Titres (Formats Courts)* [en ligne]. Consulté le 06/09/2022. Disponible à l'adresse : https://infodisc.fr/Chanson_Historique.php (cf. p. 153, 154).

LESUEUR, Daniel & DURAND, Dominic, 2022b. *InfoDisc: historique des succès musicaux en france depuis 1900* [en ligne]. Consulté le 05/09/2022. Disponible à l'adresse : https://infodisc.fr/ (cf. p. 154).

LESUEUR, Daniel & DURAND, Dominic, 2022c. *Pourquoi ce Site?* [en ligne]. Consulté le 05/09/2022. Disponible à l'adresse : https://infodisc.fr/Synthese.php (cf. p. 154).

PURE CHARTS, 2022. *A propos de PureCharts* [en ligne]. Consulté le 04/09/2022. Disponible à l'adresse : http://www.chartsinfrance.net/site/apropos.php (cf. p. 153).

SNEP, 2022a. *A propos des certifications* [en ligne]. Consulté le 02/09/2022. Disponible à l'adresse : http://snepmusique.com/les-certifications/a-propos-des-certifications/ (cf. p. 153).

SNEP, 2022b. *A propos des tops* [en ligne]. Consulté le 02/09/2022. Disponible à l'adresse : https://snepmusique.com/top-radios/a-propos-des-tops/ (cf. p. 153).

WIKIPÉDIA, 2022. *Liste des singles numéro un en France* [en ligne]. Consulté le 05/09/2022. Disponible à l'adresse : https://fr.wikipedia.org/w/index.php?title=Liste_des_singles_num%C3%A9ro_un_en_France&oldid=193890178 (cf. p. 154).

Autres références

BAUDE, Olivier, 2018. *Re: Définitions [Mail]* (cf. p. 131).

CORNEILLE, Pierre, 1643. *Cinna*. Paris : Toussaint Quinet.

DARD, Frédéric, 2010. *Si ma tante en avait*. Paris : Fleuve Noir. San-Antonio, n° 97.

JAULHAC, François, 2013. Thelma Dumas publie le 2e tome des lettres de son beau-père, sur le front, à sa belle-mère. *La Montagne* [en ligne] Consulté le 27/02/2023 (cf. p. 183).

KRISTOL, Andres, 2018. *Re: ALAVAL* (cf. p. 213).

LAMARTINE, Alphonse de, 1820. *Méditations poétiques*. 3ème édition. Paris : Dépôt de la Librairie grecque-latine-allemande.

RACINE, Jean, 1670. *Britannicus*. Paris : Claude Barbin.

RACINE, Jean, 1671. *Bérénice*. Paris : Claude Barbin.