Université Paris–Sud — Faculté des Sciences d'Orsay
École Doctorale d'Informatique de Paris–Sud
Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur

ACCENTS RÉGIONAUX EN FRANÇAIS : PERCEPTION, ANALYSE ET MODÉLISATION À PARTIR DE GRANDS CORPUS

Cécile WOEHRLING

Thèse pour le diplôme de Docteur en Sciences, spécialité Informatique
soutenue publiquement le lundi 25 mai 2009 à Orsay
devant le jury composé de

Rapporteurs	Régine ANDRÉ–OBRECHT
	Jean–François BONASTRE
Directrice	Martine ADDA–DECKER
Encadrant scientifique	Philippe BOULA DE MAREÜIL
Examinateurs	Christophe D'ALESSANDRO
	Lori LAMEL
	François PELLEGRINO

Pour Julien

Remerciements

Je tiens tout d'abord à remercier mes directeurs et directrices de thèse, Philippe Boula de Mareüil, Martine Adda-Decker et Lori Lamel pour leur conseils, leur disponibilité, leur écoute et leur gentillesse.

Je remercie également les membre de mon jury de thèse, Régine André–Obrecht, Jean–François Bonastre, Christophe d'Alessandro et François Pellegrino, pour leur participation à ma soutenance et pour leurs relectures attentives.

Je tiens à remercier Jean-Luc Gauvain, chef du groupe Traitement du Langage Parlé, pour m'avoir accueillie au sein de groupe et pour m'avoir offert la possibilité de terminer ma thèse dans de bonnes conditions.

Je remercie les membres du groupe TLP, et plus généralement les membres du LIMSI, pour les conditions agréables dans lesquelles j'ai pu effectuer mes recherches. Je pense en particulier à Daniel, Thomas, Dong et Xuan, avec qui j'ai partagé les premières années de ma thèse dans le groupe TLP ; Rena, avec qui j'ai partagé les dernières années ; Marc, François, Sylvain et Sylvain, avec qui j'ai partagé la joie de commencer une thèse en 2005 ; Gary, Guillaume, Bac, Nadi, avec qui j'ai partagé mon bureau ; Nadège, Natalie, Josep, Eric et tous ceux que je ne cite pas ici. Je tiens aussi à remercier Cécile et Hélène qui m'ont guidée et aidée lors de mes débuts dans l'enseignement à l'IUT d'Orsay.

Je remercie tout particulièrement mes collègues et amies, Bianca et Laurence, qui ont toujours été là pour partager les bons et les moins bons moments au cours de ces années.

Enfin, je remercie ma famille et mes amis pour leur soutien. Merci à Corinne, Patrick, Nicolas, Benjamin, et bien entendu Cheyenne.

Table des matières

Introduction

Notre thèse est consacrée à l'étude de variétés régionales de français. Nous nous intéressons aux accents régionaux, autrement dit aux particularités phonétiques segmentales et prosodiques qui caractérisent la prononciation d'un locuteur en fonction de son origine géographique.

Avant de poursuivre, nous devons définir ce que nous entendons par le terme *accent*. La variation dans la langue peut se manifester à plusieurs niveaux : réalisation acoustique, inventaire phonémique, lexique, prosodie… que nous n'allons pas tous prendre en compte ici. Un accent peut être défini comme *the cumulative auditory effect of those features of pronunciation, which identify where a person is from regionally or socially* [Crystal, 2003]. En adoptant cette acception du terme, nous avons exclu de nos études la variation lexicale et syntaxique pour nous focaliser sur la prononciation. Nous ferons également ici la distinction entre *accents* et *dialectes*, ces derniers correspondant à une variante orale d'une langue, avec des particularités phonétiques mais également lexicales ou syntaxiques. Nous ne traiterons donc par la suite ni des dialectes, ni de la variation diastratique (*i.e.* sociale). Nous nous concentrerons uniquement sur la variation régionale.

La variation dans la parole peut poser problème aussi bien aux humains qu'aux machines. Les humains ont une bonne capacité d'adaptation et arrivent le plus souvent à se comprendre malgré tout. Les systèmes de reconnaissance de la parole peuvent rencontrer des difficultés pour traiter de la parole avec accent : dans le cas où l'accent est inconnu et ne figure pas dans les données d'apprentisage, les systèmes sont moins performants. Du point de vue du traitement automatique, nous manquons d'informations sur les accents du français. Si de nombreuses études linguistiques ont pour objet la description précise de telle ou telle variété par rapport à un standard, nous ignorons si ces caractéristiques peuvent être retrouvées automatiquement à partir du signal de parole.

Si les performances du système de reconnaissance de la parole se dégradent en présence de parole avec accent, c'est d'abord la qualité (*i.e.* l'adéquation) des modèles acoustiques de mots qui peut porter à suspicion, dans la mesure où le vocabulaire et l'usage des mots devraient changer dans une moindre mesure. Ces changements dans la réalisation acoustique des mots gagneraient à être mieux connus. Au-delà de la motivation d'améliorer les modèles acoustiques pour rendre les systèmes de transcription plus performants face aux variations régionales et autres accents, notre objectif est aussi de rendre compte de l'utilisation combinée de grands corpus (bases d'observation des accents) et du traitement automatique pour accroître nos connaissances sur ces accents.

De grands corpus oraux comprenant des accents régionaux du français deviennent aujourd'hui disponibles : leurs données offrent une bonne base pour entreprendre l'étude des accents. Les outils de traitement automatique de la parole permettent de traiter des quantités de données plus importantes que les échantillons que peuvent examiner les experts linguistes, phonéticiens ou dialectologues.

La langue française est parlée dans de nombreux pays à travers le monde. Notre étude porte sur le français d'Europe continentale, excluant ainsi des territoires comme le Québec, l'Afrique francophone ou encore les départements d'Outre-Mer. Nous étudierons des accents régionaux de France, de Belgique et de Suisse romande.

Quelles sont les limites géographiques à l'intérieur desquelles il est possible d'affirmer que les locuteurs ont le même accent ? La réponse à cette question n'est pas évidente. Nous avons adopté la terminologie suivante, adaptée à nos données : nous parlerons d'*accent* lorsque nous ferons référence à une localisation précise telle qu'une ville ou une région donnée ; nous utiliserons le terme *variété* pour désigner un ensemble plus vaste.

Bien que de nombreuses études décrivent les particularités des accents du français, il existe moins de travaux décrivant la variation de la langue dans son ensemble, et encore moins du point de vue du traitement automatique. De nombreuses questions restent ouvertes. Combien d'accents un auditeur natif du français peut-il identifier ? Quelles performances un système automatique pourrait-il atteindre pour une tâche identique ? Les indices décrits dans la littérature linguistique comme caractéristiques de certains accents peuvent-ils être mesurés de manière automatique ? Sont-ils pertinents pour différencier des variétés de français ? Découvrirons-nous d'autres indices mesurables sur nos corpus ? Ces indices pourront-ils être mis en relation avec la perception ?

Au cours de notre thèse, nous avons abordé l'étude de variétés régionales du français du point de vue de la perception humaine aussi bien que de celui du traitement automatique de la parole. Traditionnellement, nombre d'études en linguistique se focalisent sur l'étude d'un accent précis. Le traitement automatique de la parole permet d'envisager l'étude conjointe de plusieurs variétés de français : nous avons voulu exploiter cette possibilité. Nous pourrons ainsi examiner ce qui diffère d'une variété à une autre, ce qui n'est pas possible lorsqu'une seule variété est décrite. Nous avons la chance d'avoir à notre disposition un système performant d'alignement automatique de la parole. Cet outil, qui permet de segmenter le flux sonore suivant une transcription phonémique, peut se révéler précieux pour l'étude de la variation. Le traitement automatique nous permet de prendre en considération plusieurs styles de parole et de nombreux locuteurs sur des quantités de données importantes par rapport à celles qui ont pu être utilisées dans des études linguistiques menées manuellement. Nous avons automatiquement extrait des caractéristiques du signal par différentes méthodes ; nous avons cherché à valider nos résultats sur deux corpus avec accents. Les paramètres que nous avons retenus ont permis de classifier automatiquement les locuteurs de nos deux corpus. Notre manuscrit sera organisé comme décrit ci-dessous.

Le chapitre 1 offre une vue d'ensemble de travaux existant sur l'étude des accents régionaux. Nous décrirons tout d'abord des travaux traitant des accents régionaux du français d'un point de vue linguistique et nous donnerons une description de quelques accents que nous étudierons. Nous décrirons ensuite quelques études traitant de l'iden-

tification et de la caractérisation des accents, aussi bien par l'humain que grâce à des traitements automatiques.

Le chapitre 2 présente les deux corpus sur lesquels nous avons travaillé tout au long de notre thèse. Ces deux corpus ont été automatiquement segmentés par des procédures d'alignement automatique : nous en décrirons les grandes lignes ainsi que leur application à nos données.

Le chapitre 3 donne les résultats de deux expériences portant sur la perception humaine. Nos auditeurs ont dû effectuer deux tâches : évaluer le degré d'accent et identifier l'origine géographique de locuteurs en écoutant des échantillons de parole issus de nos corpus. Les résultats ont ensuite été analysés grâce à des techniques de fouilles de données. Pour clore ce chapitre, la fusion des résultats de nos deux expériences sera présentée sous forme graphique.

Le chapitre 4 décrit les mesures de formants effectuées sur nos corpus. Nous présenterons en préambule une comparaison entre deux algorithmes d'extraction de formants. Celle-ci sera suivie d'une étude du timbre des voyelles orales à travers les valeurs des formants ; nous tracerons les triangles vocaliques correspondant à différentes variétés de français et nous discuterons des différences observées.

Le chapitre 5 est consacré à l'analyse des paramètres d'intensité, de durée et de fréquence fondamentale. Ces deux derniers paramètres sont d'abord utilisés pour caractériser certaines consonnes, en particuliers à travers des taux de voisement ; ensuite, nous y avons ajouté l'intensité dans le but d'étudier certains patrons prosodiques de début et de fin d'énoncé.

Le chapitre 6 montre une autre approche : l'exploitation de variantes dans le dictionnaire de prononciation utilisé pour l'alignement automatique. En proposant un ensemble de variantes plausibles, l'alignement automatique nous fournira des tendances sur les variantes majoritaires selon les accents. Nous pourrons dans certains cas mettre en relation ces résultats avec ceux des chapitres précédents.

Le chapitre 7 propose une validation des paramètres extraits automatiquement à travers la classification automatique des locuteurs de nos corpus. Nous essaierons de déterminer, au moyen de plusieurs classifieurs, si les paramètres que nous avons calculés aux chapitres précédents permettent de retrouver l'origine géographique d'un locuteur.

Chapitre 1

État de l'art

Comme nous nous intéressons aux accents régionaux du français, les études décrivant leurs particularités linguistiques peuvent nous servir de base pour déterminer les traits les plus intéressants, ceux qui nous permettront de trouver des paramètres pouvant être mesurés automatiquement. Nous tenterons par la suite d'exploiter ces indices pour modéliser les accents.

Les accents régionaux du français ont fait l'objet de nombreuses études linguistiques. En France, la variation diatopique (*i.e.* géographique) est d'après la littérature plus importante que la variation diastratique (*i.e.* sociale) : *(…) aux premiers mots qu'il prononce, on reconnaît un habitant de Toulouse ou de Strasbourg par rapport à un habitant de Paris, sans pouvoir toujours identifier le milieu social auquel il appartient* [Walter, 1988]. Nous présenterons tout d'abord un état de l'art des études portant sur la variation en français et nous donnerons les caractéristiques de certains accents décrits dans la littérature en section 1.1.

Les études consacrées à la caractérisation et à l'identification d'accents régionaux en français sont moins nombreuses que les travaux entrepris sur le même sujet pour d'autres langues, comme par exemple l'anglais. C'est pourquoi nous décrirons des études portant sur la caractérisation et l'identification des accents pour le français ainsi que pour d'autres langues, aussi bien par l'humain que par des techniques automatiques en section 1.2.

1.1 Les accents du français

Notre étude porte sur un sous-ensemble de toutes les variétés du français. De ce fait certains accents, comme l'accent québécois par exemple, ne font pas partie de notre travail. Dans cette section, nous donnerons une brève description des caractéristiques linguistiques de quelques accents du français et nous développerons plus en détail les accents que nous avons étudiés. La variation peut se manifester à différents niveaux, mais ainsi que nous l'avons précisé en introduction, nous avons exclu de notre étude la variation lexicale. Les caractéristiques sur lesquelles nous nous sommes concentrée et que nous décrirons sont les niveaux acoustico-phonétique, phonémique et la prosodie.

1.1.1 Les accents dans leur ensemble

Plusieurs études ont été consacrées à la description des accents régionaux du français. Les zones géographiques considérées et leur découpage diffèrent selon les travaux que nous citons ci-dessous. C'est aussi le cas du matériel employé : questionnaires, enregistrements… Malgré leurs différences, toutes ces études essaient de donner une vue d'ensemble de la variation régionale de la langue française, fondée sur des enquêtes de terrain.

Depuis le début du XX^{e} siècle, à la suite des atlas linguistiques de [Gilliéron et Edmont, 1902 1910], la variation phonétique a suscité un grand intérêt. Les atlas linguistiques permettent de distinguer les aires dialectales (au sens large) grâce au tracé d'isoglosses, frontières qui séparent des zones différant les unes des autres d'une certaine manière, par exemple par la prononciation d'un mot donné. Une tentative de visualisation de l'Atlas Linguistique de la France (ALF), qui couvre plus de 600 localités, a été entreprise plus récemment [Goebl, 2002].

L'enquête de [Martinet, 1945] est fondée sur des *Témoignages recueillis en 1941 dans un camp d'officiers prisonniers*. L'auteur a distribué des questionnaires dans lesquels les sujets devaient répondre à des questions portant sur leurs antécédents linguistiques, les systèmes des voyelles orales et nasales, les consonnes… Quelques 409 sujets ont été regroupés en 11 régions : le Midi, le Sud-Est, l'Est, le Nord, la Normandie, la Bretagne, l'Ouest, le Centre, la Bourgogne, le Centre-Nord et la région parisienne.

L'enquête de [Walter, 1982], qui prolonge la précédente, porte sur le français parlé en Europe : la France (y compris la Corse), la moitié de la Belgique, la Suisse romande et le Val d'Aoste (en Italie) sont compris dans l'étude. Les descriptions sont basées sur un corpus constitué d'enregistrements et de réponses à un questionnaire phonologique pour 111 locuteurs, regroupés en 35 variétés.

Plus récemment, [Carton *et al.*, 1983] ont accompagné leur ouvrage d'une cassette audio, permettant ainsi au lecteur d'écouter les accents du français dont il était question. Une version électronique de leurs travaux est actuellement disponible sur Internet à l'adresse accentsdefrance.free.fr.

Les études que nous venons de décrire illustrent l'évolution de la méthodologie expérimentale au cours du temps, en lien avec les instruments technologiques disponibles. Les premiers travaux utilisaient les réponses à des questionnaires papier, il a ensuite été possible d'enregistrer des locuteurs afin de pouvoir étudier les échantillons de parole et, finalement, ces échantillons ont pu être largement diffusés grâce à Internet.

1.1.2 Quelques accents en particulier

Souvent, les travaux se focalisent sur l'étude d'un accent donné, que ce soit du point de vue linguistique, sociolinguistique ou autre. Dans cette section, nous allons décrire plus en détail les accents que nous étudierons par la suite.

Le français standard

Toutes les études que nous avons mentionnées dans la section précédente ont pour objet principal les variations diatopiques (*i.e.* régionales) de la langue française. Les variantes régionales sont opposées à un français non situé géographiquement, appelé français de référence, standardisé, neutre, d'Oïl [Carton *et al.*, 1983]. Dans notre travail, nous y ferons référence sous le terme *français standard*. C'est le français véhiculé par les médias. Bien que sans référence géographique, il est souvent défini comme le français des Parisiens de milieux « bourgeois » ou intellectuels. Ce français serait aujourd'hui parlé dans la plus grande partie de la moitié nord de la France, de Rennes à Nancy [Armstrong et Boughton, 1997], à l'exception de certaines zones telles que la région Nord ou l'Alsace.

Des travaux ont été consacrés à des variétés de français incluses dans cette aire géographique. Un dialecte de Vendée a fait l'objet d'une thèse [Léonard, 1991], mais de l'aveu même de l'auteur, la francisation s'opère à grande vitesse. L'accent du Havre ne serait quant à lui qu'un mythe linguistique : des auditeurs se montrent incapables de l'identifier perceptivement [Hauchecorne et Ball, 1997].

L'inventaire phonémique du français standard peut être décrit comme suit [Fougeron et Smith, 2000] : il comporte 12 voyelles orales /a ɑ e ɛ ø œ i o ɔ u y ə/, 4 voyelles nasales /ɑ̃ ɛ̃ œ̃ ɔ̃/, 3 glides /ɥ w j/ et 17 consonnes /p b t d k g f v s z ʃ ʒ m n ɲ l ʁ/. En pratique, l'opposition /a/∼/ɑ/ tend à disparaître au profit de /a/, de même que /ɛ̃/∼/œ̃/ au profit de /ɛ̃/ [Léon, 1992]. Du point de vue prosodique, l'accent est généralement placé sur la syllabe finale.

Les accents du sud de la France

De même que pour le français standard, de nombreux termes font référence au français parlé dans le sud de la France : le français du Midi, le français méridional, le français du Sud ou encore le français d'Oc [Durand, 1995; Coquillon, 2005; Sobotta, 2006].

Les variétés méridionales de français ont été l'objet de plusieurs études. Le français de Nice, et plus particulièrement la réalisation des voyelles nasales selon les générations, a été étudié par [Thomas, 1991]. [Borrel, 1999] a, quant à lui, étudié leur réalisation à Toulouse. Les auteurs de ces deux études constatent un recul de la prononciation méridionale (voyelle dénasalisée suivi d'un appendice consonantique nasal) chez les jeunes locuteurs. [Taylor, 1996] a étudié les voyelles nasales à Aix-en-Provence. Elle observe également un recul de la prononciation méridionale, qu'elle met en lien avec l'âge et le niveau d'étude des locuteurs.

La thèse de [Sobotta, 2006] est en partie consacrée à l'étude de locuteurs aveyronnais. L'auteur déplore l'absence d'étude sur le français méridional dans son ensemble : il lui est de ce fait difficile de positionner les observations faites sur les Aveyronnais par rapport à l'ensemble des Méridionaux. Elle indique toutefois que les différences entre les divers français du Midi ne semblent pas être trop importantes.

L'accent de Marseille a fait l'objet de plusieurs études. [Binisti et Gasquet-Cyrus, 2003] l'ont étudié à partir d'enregistrements et de questionnaires. Ils opposent un stéréotype na-

tional de l'accent marseillais à trois accents émergeant de leurs données : l'accent des « vrais Marseillais », l'accent de la « bourgeoisie marseillaise » et l'accent des « quartiers Nord ». Les particularités prosodiques de l'accent marseillais ont fait l'objet d'une thèse [Coquillon, 2005]. L'auteur a observé plus de variation de fréquence fondamentale ainsi qu'un contour mélodique particulier en « chapeau mou » [1] chez des locuteurs marseillais comparés à des locuteurs du français standard. Si elle mentionne également une différence entre Marseillais et Toulousains au niveau de l'analyse rythmique du schwa final, elle indique également que les particularités de l'accent marseillais sont, pour beaucoup, partagées avec d'autres parlers du sud de la France.

Toutes ces études mentionnent des traits caractéristiques de l'accent dont elles traitent. Nous présentons ci-dessous quelques-uns des traits qui sont partagés par les accents du sud et qui les distinguent du français standard :
– les voyelles nasales sont souvent partiellement, voire totalement dénasalisées et suivies d'un appendice consonantique nasal ;
– la réalisation de nombreux schwas (ou *e* muets) [Durand *et al.*, 1987] ;
– la réduction des oppositions semi-ouvert/semi-fermé pour les voyelles moyennes, dont la distribution suivrait la loi de position [Durand, 1995] ;
– la simplification de groupes consonnantiques complexes ;
– une prosodie différente de celle du français standard (sans mesures précises).

Le français en Alsace

L'Alsace n'est traditionnellement pas incluse dans la grande moitié nord de la France où le français standard est parlé. Cette région a un passé linguistique complexe [Bister-Broosen, 2002] : la langue officielle y a successivement été le français et l'allemand. Il faut également prendre en compte le dialecte alsacien (dialecte germanique), lequel influence la prononciation de ses locuteurs : en 1998, un locuteur sur deux déclare encore le parler couramment [Bothorel-Witz, 2000]. [Vajta, 2002] a mené en Alsace une enquête linguistique dont le but était d'observer le recul des langues germaniques par rapport au français. Les mesures étaient entre autres effectuées sur des marqueurs tels que la diphtonguaison, ou encore l'accentuation initiale. La prosodie semble jouer un rôle important dans cet accent, mais les études sur le sujet manquent : nous n'avons pu trouver en la matière qu'une étude comparant les variétés du nord et de l'est (Lorraine) de la France d'un point de vue prosodique [Carton *et al.*, 1991].

Parmi les traits caractérisant l'accent alsacien, nous pouvons citer :
– l'aspiration des consonnes plosives sourdes /p t k/ ;
– une opposition de voisement peu ou pas marquée pour certaines consonnes (par exemple /p/∼/b/) ;
– une prosodie différente de celle du français standard.

1. Le contour en « chapeau mou » est caractérisé par une montée de F_0 suivie par un plateau plus ou moins long, et se termine par une chute d'une hauteur équivalent à la montée.

Le français en Belgique

La Belgique possède trois langues officielles : le flamand, proche du néerlandais, le français et l'allemand. Le français est la langue officielle de la région wallone (partie sud de la Belgique), tandis que la région de Bruxelles est bilingue français–néerlandais.

Selon [Hambye et Francard, 2004], deux thèses antagonistes coexistent actuellement : d'après la première, le français parlé en Belgique représenterait une variété spécifique ; la deuxième, au contraire, soutient que le français de Belgique serait similaire au français standard. Il n'est pas facile de se prononcer en faveur de l'une ou de l'autre de ces hypothèses : les études systématiques sont rares. De plus, certains travaux confondent les variations diatopiques (*i.e.* régionales) et diastratiques (*i.e.* sociales). Une autre question reste ouverte : existe-t-il des traits communs à l'ensemble de la Belgique (incluant les locuteurs wallons et bruxellois) ?

D'après [Francard, 2000], l'accent belge est identifié à l'étranger comme celui des types bruxellois populaires, or cet accent n'est plus observable aujourd'hui en dehors de situations volontairement humoristiques. Pour [Pohl, 1983], *celui qui franchit la frontière franco-belge, en n'importe quel point, remarque une différence d'« accent ». (…) Le Parisien ou l'habitant des départements du nord qualifient de « belge » assez indistinctement cet « accent » qui les surprend au premier abord*. Le même auteur examine la situation symétrique : *Pour (le Belge), à vrai dire, trois « accents se partagent en gros la France : l'« accent français » proprement dit (…), l'« accent du Midi » et l'« accent alsacien »*.

Outre ces considérations, il existe quelques études décrivant des traits des accents de Belgique. L'accent liégeois est connu pour être l'un des plus typiques et des plus résistants de la région wallonne [Hambye et Simon, 2004]. Ces auteurs ont étudié l'allongement des voyelles en position pénultième et finale, trait caractéristique du français de Belgique, et ce pour 4 locuteurs de la ville de Liège. [Pohl, 1983] a étudié la phonologie du français parlé en Belgique à travers les réponses de 13 locuteurs à des questionnaires phonologiques, adaptés de ceux de Martinet (section 1.1.1). Il en ressort un certain nombre de traits qui, d'après l'auteur, pourraient se retrouver en France, mais avec une fréquence ou une constance moindre. Nous pouvons citer les traits suivants :

- l'opposition /ɛ̃/–/œ̃/ est maintenue, donnant un système vocalique à quatre voyelles nasales ;
- les oppositions de quantité vocalique sont utilisées pour différencier /a/ et /ɑ/ ou encore pour marquer le féminin (*ami* [ami] ∼ *amie* [ami ː]) ;
- les oppositions /o/∼/ɔ/ et /e/∼/ɛ/ sont maintenues en finale absolue (*peau* [po] ∼ *pot* [pɔ]) ;
- les consonnes sonores finales ont tendance à s'assourdir ;
- la diérèse est parfois faite là où en français standard il y a généralement une synérèse (*lion* [liõ]) ;
- une certaine nasalisation régressive de /ɛ/ s'observe (par exemple dans *laine*) ;
- la prosodie est différente de celle du français standard et se caractérise par l'allongement de certaines voyelles.

Le français en Suisse

La Suisse possède trois langues officielles : l'allemand, l'italien et le français. Ce dernier est parlé dans la partie ouest du pays, la Suisse romande. Si le français parlé en Suisse romande partage l'essentiel de ses traits avec le français standard [Singy, 1995 1996], il présente toutefois certaines particularités. Un grand nombre d'entre elles est dû au substrat dialectal francoprovençal [Singy, 2000]. La plupart des différences sont d'ordre lexical ; nous ne les décrirons pas ici, étant donné qu'elles sortent du cadre de nos travaux. D'après [Métral, 1977], il est difficile de trouver des traits permettant d'identifier le français de Suisse romande dans son ensemble : certains traits vont au delà des frontières étatiques, tandis que d'autres sont propres à un canton donné, ou diffèrent selon la situation géographique (montagne, vallée...).

Le français parlé dans le canton de Vaud a fait l'objet d'études sociolinguistiques [Singy, 1995 1996; Singy, 2000]. Le système vocalique des locuteurs suisses a également été étudié [Métral, 1977]. [Grosjean *et al.*, 2007] ont plus précisément focalisé leur travail sur les variations de durée des voyelles. Ils ont montré que, sur 25 participants, les locuteurs suisses du canton de Neuchâtel produisent des voyelles de longueur significativement différentes, au contraire des locuteurs parisiens. Les Suisses perçoivent également ces oppositions de longueur, contrairement aux Parisiens. Le timbre du schwa de locuteurs suisses a été étudié par [Bürki *et al.*, 2008], en comparaison avec des locuteurs du français standard et du Québec. L'étude est toutefois limitée à un peu moins de 300 schwas (réalisés ou non) au total. Elle conclut à un schwa plus ouvert en français de Suisse qu'en français standard.

Les caractéristiques d'ordre phonétique et prosodique les plus remarquées pour les locuteurs suisses romands sont les suivantes :
- l'opposition /ɛ̃/–/œ̃/ est maintenue, donnant un système vocalique à quatre voyelles nasales ;
- les oppositions de quantité vocalique sont maintenues ;
- la prosodie se distingue de celle du français standard.

Il est intéressant de noter que certaines des particularités mentionnées ci-dessus sont communes aux accents de Suisse et de Belgique. Selon certains stéréotypes [Singy, 2000], il faudrait également ajouter la caractéristique suivante : les locuteurs suisses présentent un débit plus lent que d'autres, notamment les locuteur du français standard. Toutefois, toutes les études ne valident pas cette différence de débit [Schwab *et al.*, 2008].

1.2 Caractérisation et identification d'accents

Les accents, ou les variétés d'une langue, peuvent être étudiés de différente manières. Dans nos travaux, nous adopterons deux approches : la perception humaine et le traitement automatique. Ces deux approches ne s'appliquent pas sur les mêmes quantités de données. Un test perceptif ne peut mettre en jeu que peu de données, contrairement au traitement automatique, idéal pour traiter de grands corpus. Ces deux approches peuvent se révéler complémentaires : l'étude des performances et des capacités des humains mesurées par des tests perceptifs fournit une référence lors de l'évaluation des performances du traitement au-

tomatique. Les indices utilisés par les auditeurs humains peuvent également nous indiquer les paramètres qui permettraient d'identifier automatiquement tel ou tel accent.

Nous aborderons successivement ces deux aspects en présentant d'abord des travaux portant sur la perception humaine, avant de nous intéresser aux approches automatiques.

1.2.1 Approches perceptives

Les réactions que des sujets ont à l'écoute d'échantillons de parole réels peuvent être différentes de celles qu'ils pensaient avoir. Une personne peut par exemple se déclarer capable d'identifier un locuteur d'une certaine variété de français, et ne pas y parvenir une fois confrontée à des extraits audio. En sociolinguistique, les études se focalisent souvent sur les représentations plus ou moins stéréotypées qu'un sujet se fait de variétés spécifiques. Ces représentations peuvent se révéler différentes des comportements en réaction à l'écoute d'échantillons de parole, lorsque des auditeurs doivent évaluer des locuteurs, les juger subjectivement en termes d'intelligence, d'amabilité, de virilité, de correction, etc. [Preston, 1993; Hintze *et al.*, 2000]. Mais la plupart des études sont descriptives et ne permettent pas de prédire avec certitude les caractéristiques les plus discriminantes qui sont associées à un accent donné.

La méthode dite « des flèches » consiste à demander à des sujets de citer les lieux dont ils se sentent proches par la manière de parler et ceux dont ils pensent être complètement éloignés [Preston, 1989]. À partir des réponses obtenues, il est possible de construire une carte, par exemple en reliant par des flèches les points désignés comme proches. Cependant, cette méthode ne fonctionne que pour des endroits géographiquement proches, et des sujets issus de ces régions : pour des points éloignés, il est impossible de dessiner les flèches et, de ce fait, de refléter leur proximité éventuelle dans la façon de parler.

La perception humaine d'accents régionaux a fait l'objet d'études dans différentes langues. Dans leurs travaux, [Clopper et Pisoni, 2004] ont montré que, sans entraînement préalable ni retour (*feedback*), des auditeurs américains, invités à écouter des compatriotes de différents accents et à localiser leur origine géographique sur une carte des États-Unis, sont capables de distinguer trois grandes régions : Nouvelle Angleterre, Sud et Nord-Ouest. Une thèse sur les dialectes norvégiens et néerlandais [Heeringa, 2004] a établi un parallèle entre une étude perceptive d'une part et les distances entre les transcriptions phonétiques ou les réalisations acoustiques de différents mots d'autre part. Des travaux ont été consacrés à l'identification de quatre variétés de néerlandais et de cinq variétés d'anglais par des auditeurs des pays concernés [van Bezooijen et Gooskens, 1999]. Plusieurs niveaux de réponses étaient proposés : les résultats montrent une très bonne identification du pays d'origine des locuteurs (environ 90 %) et des scores moins élevés pour l'identification de l'origine exacte des locuteurs (40 à 50 %). Les travaux de [Williams *et al.*, 1999] portent sur l'identification de six variétés de gallois par des auditeurs des régions concernées. Le taux d'identification est plutôt bas pour les jeunes auditeurs (20 à 44 %), meilleur pour des auditeurs plus âgés et exerçant la profession d'enseignants (jusqu'à 85 %). Une autre expérience de perception, portant sur les régions germanophones, a montré que les dialectes suisses alémaniques, autrichiens et saxons étaient les mieux identifiés [Burger et Draxler, 1998].

Concernant le français, les études empiriques auxquelles nous avons pu nous reporter n'impliquent que deux ou trois variétés. Celle de [Bauvois, 1996] porte sur des variétés de français parlé en Belgique et dans différents pays d'Afrique. [Armstrong et Boughton, 1997] examinent la perception du français parlé à Nancy et à Rennes, deux villes appartenant historiquement au domaine d'oïl, quasiment symétriques par rapport à Paris : il en ressort que la classe sociale des locuteurs est bien identifiée, mais pas leur provenance géographique. De même [Hauchecorne et Ball, 1997] concluent que « l'accent du Havre » est plus un accent social à l'image négative et présent en d'autres lieux qu'une réalité géolinguistique identifiable. Le chapitre de la thèse de [Sobotta, 2006] qui traite de la perception discute la question de la gradience de la variation diatopique (ou géographique) et diaphasique (ou stylistique). Ses expériences, séparées en deux parties, portent sur 5 groupes de témoins : Méridionaux de l'Aveyron, Guadeloupéens, Aveyronnais et Guadeloupéens ayant migré à Paris et Parisiens ; elles montrent une perception continue des accents de ces locuteurs par des auditeurs de région parisienne et orléanaise. Une autre épreuve d'identification sur le français et le francique parlés dans des régions frontalières de France, Belgique et Luxembourg a montré que des auditeurs bilingues de ces trois mêmes régions se montrent capables de reconnaître l'origine géographique des locuteurs (française, belge ou luxembourgeoise), en français plus encore qu'en francique [Rispail et Moreau, 2004]. Dans l'ensemble cependant, la variation sous l'angle des aspects phonétiques a donné lieu à beaucoup moins de travaux que le contact de langues (alternances codiques, emprunts, etc.).

1.2.2 Approches automatiques

Les approches automatiques que nous avons pu trouver dans la littérature ont des objectifs différents. Certaines études s'attachent à décrire les caractéristiques des variétés de langues ; d'autres proposent de classifier automatiquement les accents. Certaines études abordent ces deux points, ou font un parallèle avec des études perceptives. Enfin, il existe également des travaux portant sur la dégradation de la reconnaissance automatique de la parole, ou ASR (*Automatic Speech Recognition*).

En français, peu d'études traitent des accents régionaux. Des travaux ont été consacrés aux accents étrangers en français [Boula de Mareüil *et al.*, 2004; Vieru-Dimulescu, 2008]. [Bartkova et Jouvet, 2004] ont montré l'interêt d'une modélisation multilingue pour rendre compte des prononciations non-standard de locuteurs non-natifs du français. Les phénomènes de schwa et de liaison ont été étudiés sur le français à travers des variantes de prononciation [Adda-Decker *et al.*, 1999], mais pas dans le but de mettre en évidence des variations régionales. Les accents régionaux ont en revanche fait l'objet d'études dans d'autres langues.

Les systèmes de reconnaissance de la parole peuvent voir leurs performances dégradées lorsqu'ils doivent traiter de la parole avec accent, quand les données sont éloignées de celles, plutôt standard, qui ont servi à entraîner le système. [Yan et Vaseghi, 2002] ont mesuré la dégradation des performances d'un système de reconnaissance de la parole avec des données d'anglais américain et britannique. Ils ont observé que le taux d'erreur augmente lorsque le système est entraîné et testé sur des variétés différentes. Les auteurs ont également montré que les accents de leurs données présentent des différences tant au niveau phoné-

tique qu'au niveau prosodique. Même si la tâche est plus complexe lorqu'il s'agit d'accents étrangers, les dialectes et les accents régionaux posent des problèmes importants. D'après [Van Compernolle, 2001], les taux d'erreurs doublent voir triplent lorsque le système est entraîné du néerlandais et testé sur du flamand et inversement. Les mêmes problèmes se posent pour l'anglais britannique et américain. Pour prendre en compte ces variation, plusieurs approches coexistent : l'ajout de variantes de prononciation dans les dictionnaires de prononciation et l'adaptation de modèles acoustiques à partir de corpus avec accents.

Plusieurs études se sont concentrées sur l'estimation des distances entre variétés de langues. La méthodologie appliquée est similaire : un *clustering* des accents considérés est construit à partir d'une mesure de distance appliquée à des paramètres issus du signal de parole. Si le clustering permet d'observer quelles variétés sont plus ou moins proches les unes des autres, il n'est pas exactement comparable à une technique de classification. Ces travaux sont plutôt dédiés à la description de la variation. Les accents régionaux du suédois ont ainsi été étudiés à partir d'enregistrements téléphoniques d'environ 5000 locuteurs [Salvi, 2003]. Dix-neuf variétés de suédois ont été considérées ; les paramètres acoustiques utilisés sont les MFCC [2] (*Mel-frequency cepstral coefficients*) et l'énergie. Les distances entre variétés sont calculées grâce à la distance de Bhattacharyya [3]. Les regroupements obtenus par clustering hiérarchique semblent cohérents avec les connaissances linguistiques.

Une thèse sur les dialectes norvégiens et néerlandais [Heeringa, 2004] a également développé une cartographie des distances phonologiques et acoustiques qui existent au niveau lexical au sein d'un même ensemble dialectal. L'auteur a mesuré les distances de Levenshtein [4] entre les transcriptions phonétiques ou les réalisations acoustiques de différents mots. Le travail sur le norvégien a été poursuivi [Heeringa *et al.*, 2009] en utilisant uniquement des mesures acoustiques (formants et taux de passage par zéro, ou *zero crossing rate*) sans aucune information liée à une transcription phonétique, ce qui n'était pas le cas dans le travail précédent. Sur le néerlandais encore, [ten Bosch, 2000] a mis en relation les distances phonologiques calculées précédemment avec des distances calculées à partir de la reconnaissance automatique de la parole. Ces distances étaient fondées sur la dégradation des performances du système, en considérant un système entraîné en laissant de côté une région.

Plusieurs études ont porté sur les accents de l'anglais. [Ghorshi *et al.*, 2007] ont travaillé sur l'anglais parlé en Grande-Bretagne, en Australie et aux États-Unis. Une mesure de *cross-entropy* [5] a été appliquée sur les formants et les mesures cepstrales, afin de modéliser les effets de l'accent sur les unités phonétiques de la parole. Les auteurs concluent que les cepstres portent moins d'informations que les formants pour modéliser les accents de l'anglais.

Dans la même veine, une étude a été consacrée à la classification d'anglais britannique et américain [Hansen *et al.*, 2004]. Les auteurs ont également étudié la structure prosodique, toujours dans le but d'améliorer les scores de la reconnaissance automatique de la parole.

2. Les coefficients MFCC sont extraits du cepstre (transformation du spectre du signal de parole). Ils sont sensés représenter la forme du conduit vocal du locuteur.

3. La distance de Bhattacharyya permet de calculer une distance entre deux distributions en prenant en compte leurs deux variances.

4. La distance de Levenshtein permet de calculer une distance d'édition séparant des chaînes de caractères.

5. La *cross-entropy* est une mesure de la différence entre deux distributions.

La métrique ACCDIST (*Accent Characterisation by Comparison of Distances in the Intersegment Similarity Table*) a été proposée par [Huckvale, 2004] dans le but de quantifier la similarité entre des accents. Le principe de cette mesure consiste à s'appuyer sur les relations entre les réalisations de certains segments plutôt qu'à leurs valeurs intrinsèques, afin de prendre en compte les particularités du système de prononciation du locuteur plutôt que celles de sa voix. En pratique, il s'agit de calculer une table de similarité entre segments pour caractériser chaque locuteur ; la distance entre deux locuteurs est donnée par le calcul de la corrélation entre deux tables. La mesure ACCDIST a été appliquée aux accents régionaux de l'anglais britannique [Huckvale, 2007]. L'étude a été conduite sur une vingtaine de phrases de 20 locuteurs issus de 14 régions du corpus ABI (*Accents of the British Isles*, [D'Arcy *et al.*, 2005] – parole lue). Les paramètres acoustiques (formants et enveloppe spectrale) ont été calculés sur les voyelles. La mesure ACCDIST a été comparée à d'autres distances (par exemple la distance euclidienne) sur une tâche de clustering hiérarchique : ACCDIST montre de meilleures performances dans la mesure où elle regroupe des locuteurs censés avoir le même accent et qu'elle montre des regroupements plus facilement intérpretables à la lumière de connaissances linguistiques.

L'étude des accents du corpus ABI avec la mesure ACCDIST a été poursuivie dans la thèse de [Ferragne, 2008]. Les données (parole lue) sont issues de 13 régions du corpus ABI. Les voyelles, caractérisées par les MFCC, ont servi à classifier les locuteurs grâce à la méthode du plus proche voisin (par validation croisée, distances calculées avec ACCDIST). La thèse inclut également une analyse des voyelles (à travers des mesures de formants) et du rythme (à travers des proéminences de durée et d'intensité) mais ces mesures n'ont pas été utilisées pour la classification.

1.3 Conclusion

Les méthodes employées pour décrire la variation de la langue ont, nous l'avons vu, évolué au cours du temps. Les avancées technologiques ont permis d'aller du recueil de données par questionnaires sur papier à l'étude d'enregistrements. Aujourd'hui, les technologies ont encore évolué : les quantités de données disponibles sont plus importantes, et le traitement automatique de la parole nous autorise à analyser ces grands corpus.

Les variétés de français ont été décrites du point de vue linguistique : les caratéristiques mentionnées dans ces travaux peuvent-elles être retrouvées grâce à une analyse automatique ? Nous avons constaté que, parmi les travaux concernant la caractérisation ou l'identification automatique d'accents, bien peu portent sur le français. Il existe pourtant des corpus d'accents régionaux en français qu'il est possible d'étudier : c'est ce qui nous a amenée à entreprendre les travaux qui font l'objet de cette thèse.

Nous appliquerons à nos données les deux approches que nous avons mentionnées dans ce chapitre. Nous étudierons la perception à travers des tests d'identification de variétés régionales de français, à la manière de [Clopper et Pisoni, 2004]. Nous appliquerons diverses approches automatiques : nous mesurerons certains paramètres acoustiques, et nous utiliserons des variantes de prononciation adaptées aux accents régionaux. Nous tenterons également de classifier certaines variétés de français, comme cela a été réalisé dans d'autres

langues. Nous présenterons au chapitre suivant les corpus sur lesquels nous travaillons ainsi que l'alignement automatique en phonèmes.

Chapitre 2

Corpus et alignement automatique

Nous avons aujourd'hui à notre disposition d'importants corpus audio incluant des accents régionaux du français. Pour analyser de telles quantités de données, nous pouvons nous aider de l'alignement automatique en phonèmes dérivé du système de reconnaissance de la parole mis au point au LIMSI [Gauvain *et al.*, 2005]. Grâce à la segmentation, de nombreuses possibilités d'étude s'offrent à nous.

Nos études s'appuient sur deux corpus. Le premier, issu du projet « Phonologie du Français Contemporain » (PFC), a été constitué en suivant un protocole bien établi — décrit ci-dessous — et met en jeu des locuteurs aux origines relativement contrôlées. Il offre également la possibilité d'étudier plusieurs « styles » de parole, comme la lecture ou l'entretien. Le second corpus est constitué de conversations téléphoniques (*Conversational Telephone Speech*, CTS) faisant intervenir des locuteurs aux origines moins contrôlées. Nos deux corpus ont été automatiquement segmentés en phonèmes grâce au système développé au LIMSI. Après une présentation du corpus PFC (section 2.1) puis du corpus CTS (section 2.2), nous décrirons la procédure d'alignement automatique en phonèmes ainsi que son application en section 2.3.

2.1 Le corpus PFC

2.1.1 Le projet PFC

Le projet PFC [Delais-Roussarie et Durand, 2003; Durand *et al.*, 2002; Durand *et al.*, 2005] (www.projet-pfc.net) est un projet international codirigé par Jacques Durand (ERSS, Université de Toulouse-Le Mirail), Bernard Laks (MoDyCo, Université de Paris X) et Chantal Lyche (Universités d'Oslo et de Tromsø). Il s'inscrit dans le sillage des travaux de [Martinet, 1945] et de la grande enquête phonologique de [Walter, 1982], qui portait sur une trentaine de régions francophones d'Europe. Le projet PFC offre une base de données de français oral contemporain dans l'espace francophone, établie d'après un protocole commun. Le site indique actuellement 74 enquêtes, dont 24 sont terminées.

Des enregistrements d'informateurs bien ancrés géographiquement — étant nés et ayant passé une grande partie de leur vie en un même lieu — sont collectés en différents points de la Francophonie. Dans chacun de ces points d'enquête, le même protocole est soumis à une dizaine de locuteurs, comptant un nombre équivalent d'hommes et de femmes répartis en au moins deux tranches d'âge. Pour chaque locuteur, plusieurs styles de parole sont représentés, dans la lignée labovienne [Labov, 1972] : la lecture d'une liste de mots, d'un texte d'une vingtaine de phrases (voir annexe A, p. 143) ainsi que de la parole spontanée sous forme d'un entretien guidé (plus formel) et d'une conversation libre (moins formelle). Les enregistrements de parole spontanée sont de durée variable, une dizaine de minutes environ étant au total transcrite orthographiquement pour chaque locuteur. Une fiche d'information, comportant au minimum l'âge et le sexe est également disponible pour chacun.

2.1.2 Notre corpus

Le corpus que nous avons analysé est un sous-ensemble du corpus décrit ci-dessus : nous avons progressivement inclus 16 points d'enquête parmi ceux qui étaient disponibles au moment de nos études. Nous n'avons pas conservé les enregistrements de la lecture de la liste de mots, qui représente un style de parole très particulier et peu utilisé en traitement automatique. Nous présentons ici nos données par point d'enquête et regroupées en cinq variétés de français.

FIGURE 2.1. Localisation des 16 points d'enquête analysés

Nous avons regroupé nos points d'enquête afin de pouvoir étudier des régions plus étendues : il aurait en effet été plus difficile d'essayer de caractériser chacun des 16 points. Les regroupements que nous avons effectués peuvent être discutés et nous aurions peut-être pu trouver des paramètres discriminant certains points que nous avons regroupés au sein de la même variété, mais cela n'a pas été le cas avec les paramètres que nous avons étudiés par la suite. Nous avons par exemple considéré une variété standard : nous entendons par là que

les points d'enquête qui la composent sont, par rapport à d'autres, proches de ce que l'on pourrait appeler standard, et partagent un certain nombre de caractéristiques entre eux.

Nous avons 6 points d'enquête situés dans la moitié nord de la France, considérés comme français standard (Brécey, Brunoy, Dijon, Lyon Villeurbanne, Roanne, et Treize-Vents), 5 dans le sud de la France (Biarritz, Douzens, Lacaune, Marseille, et Rodez), 1 en Alsace (Boersch), 3 en Belgique (Tournai, Gembloux et Liège) et 1 en Suisse romande (Nyon). Leur emplacement géographique est indiqué sur la figure 2.1.

	#loc.	#H+F	âge moyen	durée texte (min)	durée entretien guidé (min)	durée conversation libre (min)	durée totale (h)
Brécey	11	5+6	44	28	105	60	3
Brunoy	10	5+5	55	26	54	54	2
Dijon	8	3+5	34	18	80	72	3
Lyon	11	5+6	38	23	60	56	2
Roanne	9	5+4	66	23	47	67	2
Treize-Vents	8	5+3	46	19	73	65	3
Total Standard	**57**	**28+29**	**47**	**137**	**419**	**374**	**16**
Biarritz	12	5+7	47	30	111	58	3
Douzens	10	5+5	50	27	112	64	3
Lacaune	13	5+8	56	43	64	64	3
Marseille	8	5+3	54	23	0	0	0,5
Rodez	9	4+5	46	20	84	66	3
Total Sud	**52**	**24+28**	**51**	**143**	**371**	**252**	**13**
Boersch	**13**	**6+7**	**61**	**32**	**177**	**0**	**4**
Gembloux	12	6+6	44	31	98	68	3
Liège	12	6+6	44	32	63	61	3
Tournai	12	6+6	41	30	64	78	3
Total Belgique	**36**	**18+18**	**43**	**93**	**225**	**207**	**9**
Nyon	**12**	**7+5**	**46**	**31**	**126**	**60**	**4**
total	**170**	**83+87**	**48**	**436**	**1318**	**893**	**44**

TABLEAU 2.1. Description du corpus PFC par point d'enquête : nombre de locuteurs (total, hommes et femmes), âge moyen, quantité de parole transcrite pour les 3 styles retenus.

L'ensemble est constitué de plus de 150 locuteurs, environ autant d'hommes que de femmes appartenant à différentes catégories d'âge. Nous avons travaillé sur les données transcrites orthographiquement : alignées en phonèmes, elles représentent au total des dizaines d'heures de parole (plus de 44). Le Tableau 2.1 donne le nombre de locuteurs et la quantité de parole transcrite par point d'enquête, ainsi que pour chacune des variétés étudiées par la suite. Les variétés standard et Sud regroupent un nombre plus élevé de points d'enquête et représentent de ce fait une quantité de données relativement importante. La Belgique est représentée par trois points. Nous n'avons respectivement qu'un seul point d'enquête en

Alsace et en Suisse romande. Les conclusions que nous pourront tirer par la suite sur les caractéristiques de ces deux régions ne seront validées que pour les deux villes concernées. Il pourra être intéressant de les généraliser lorsque d'autres données seront disponibles dans la base PFC. Il faut également noter que certains points d'enquête sont incomplets : nous n'avons que le texte pour Marseille [1], et nous n'avons pas d'entretien libre pour Boersch (cet entretien s'est déroulé en alsacien dans la majorité des cas).

2.2 Le corpus CTS

Le corpus CTS (*Conversational Telephone Speech*), est constitué de conversations téléphoniques entre des locuteurs appelant de 7 grandes régions françaises (Centre, Est, Nord, Ouest, Paris, Sud-Est et Sud-Ouest). Il s'agit d'un corpus interne au LIMSI, collecté dans le cadre de différents contrats de recherche avec des industriels. Le sexe des locuteurs est connu ; nous avons de plus à notre disposition un degré d'accent évalué par les transcripteurs sur une échelle discrète allant de 1 à 3. Nous n'avons par contre aucune information concernant l'âge, le lieu exact de résidence ou encore l'histoire linguistique de ces locuteurs, hormis les informations que nous pouvons glaner en écoutant les conversations elles-mêmes. Moins contrôlé que le corpus PFC, et de qualité audio différente (téléphonique), le corpus CTS nous permettra de vérifier si les caractéristiques observées sur le premier restent valables sur le second.

	#demi conversations (H+F)	durée totale (heures)	durée moyenne conversation (min)
Standard	101+189	35	14
Sud	77+143	26	14
Alsace	18+22	4	13
total	206+384	65	14

TABLEAU 2.2. Description du corpus CTS : nombre de conversations, durée totale et moyenne des enregistrements.

Nous avons regroupé les régions de départ pour les faire correspondre aux groupes étudiés et disponibles dans le corpus PFC. Nous avons ainsi constitué trois groupes d'enregistrements correspondant au français standard (Centre, Ouest, Paris), au sud de la France (Sud-Est et Sud-Ouest) et à l'Alsace (une partie de l'Est, l'origine des locuteur retenus a été déterminée en écoutant les enregistrements). Au total, nous avons à notre disposition un peu moins de 300 conversations d'environ 14 minutes chacune, soit environ 65 heures de parole. Le détail par groupe est donné Tableau 2.2. Étant donné qu'un même locuteur peut participer à plusieurs appels, nous avons préféré compter des « demi conversations » plutôt que des locuteurs. Tout comme pour le corpus PFC, les données sont inégalement réparties entre les variétés standard et Sud (qui totalisent un nombre d'heures de parole relativement élevé) et l'Alsace, pour laquelle nous n'avons que quatre heures de parole.

1. La parole spontanée, à travers de courts extraits, n'a été utilisée que dans les premiers tests perceptifs présentés dans le chapitre suivant.

2.3 Alignement automatique en phonèmes

Nos deux corpus, transcrits orthographiquement, ont été alignés en phonèmes à l'aide de l'alignement automatique dérivé du système de reconnaissance de la parole du LIMSI [Gauvain *et al.*, 2005; Lamel et Gauvain, 2003]. Après une description de ce système dans la section suivante, nous donnerons en illustration certains paramètres que l'alignement nous a permis de calculer.

2.3.1 Procédure d'alignement en phonèmes

La procédure d'alignement en phonèmes permet d'obtenir une segmentation en phonèmes à partir d'un signal de parole et de sa transcription orthographique. Elle nécessite d'une part un dictionnaire de prononciation et d'autre part des modèles acoustiques. La méthode, dont le fonctionnement est illustré par la figure 2.2, a été utilisée dans plusieurs études qui ont permis de la valider [Adda-Decker et Lamel, 1999; Adda-Decker *et al.*, 2005; Gendrot et Adda-Decker, 2005].

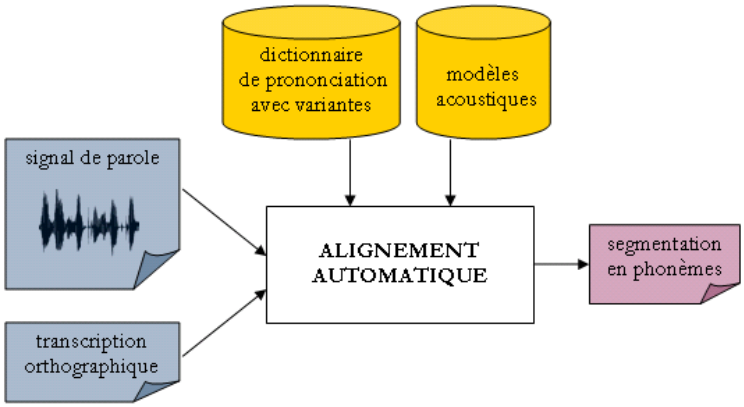


FIGURE 2.2. Alignement automatique en phonèmes.

Le dictionnaire de prononciation permet d'associer une ou plusieurs formes phonémiques à chaque entrée orthographique qu'il contient (ex. *côte* [kot, kɔt, kotə, kɔtə]). Il est possible de proposer des variantes plus ou moins « standard » : réalisation optionnelle du schwa, degré d'ouverture des voyelles, variantes dans la réalisation des voyelles nasales... Un alignement forcé, utilisant un dictionnaire « standard » avec peu de variantes de prononciation, donnera une transcription plutôt phonémique, qui permettra d'étudier la variation à travers des mesures acoustiques (ex. la variation des formants du /o/ de *côte*). Le dictionnaire « standard » que nous avons employé comporte des variantes telles que certaines liaisons, ou encore la réalisation optionnelle de certains schwas. Un alignement utilisant un dictionnaire proposant des variantes de prononciations régionales permet d'obtenir une transcription plus phonétique, du moins pour les phonèmes concernés par les variantes proposées. La variation apparaît dans ce cas dans la transcription, à travers les variantes choisies par le système (ex. *côte* correspond-il davantage au modèle acoustique de [kot] ou de [kɔt] ?). Nous adopterons ces deux approches dans la suite de notre travail.

LIMSI	i	e	E	y	@	x	a	c	o	u	I	A	O		h	w	j
API	i	e	ɛ	y	ø	ə	a	ɔ	o	u	ɛ̃	ɑ̃	ɔ̃		ɥ	w	j
LIMSI	p	b	t	d	k	g	f	v	s	z	S	Z	m	n	N	l	r
API	p	b	t	d	k	g	f	v	s	z	ʃ	ʒ	m	n	ɲ	l	ʁ
LIMSI	. silence			H respiration			& hésitation										

TABLEAU 2.3. Inventaire des unités du système d'alignement automatique du LIMSI, avec les symboles LIMSI et API.

Nos modèles acoustiques pour le français comprennent 33 phonèmes plus 3 modèles pour le silence, les hésitations et les respirations. Les modèles acoustiques confondent le schwa et le /œ/ de *leur*. Ce dernier symbole pourra apparaître dans les triangles vocaliques car l'orthographe permet de repérer les segments acoustiques correspondant à ce phonème. Il est cependant beaucoup moins fréquent que le /ə/. Le tableau 2.3 dresse l'inventaire de ces 36 unités et donne la correspondance dans l'Alphabet Phonétique International (API).

Les modèles acoustiques correspondent à un modèle de Markov caché à trois états modélisant chaque unité (ou phone) composant le signal de parole. Chaque modèle est obtenu par l'entraînement sur des données audio d'apprentissage. Dans notre cas, les modèles ont été entraînés sur des données différentes selon la bande passante de l'enregistrement : bande large ou bande téléphonique. Les modèles de phones peuvent tenir compte du ou des phones qui le précèdent et le suivent ; le modèle est alors dit dépendant du contexte. Ce n'est pas le cas pour nos alignements : les modèles que nous avons utilisés sont indépendants du contexte pour mieux saisir les caractéristiques générales propres à chaque phonème.

2.3.2 Quelques mesures issues de l'alignement

Dans cette section, nous nous limiterons au calcul de grandes caractéristiques du corpus à partir d'un dictionnaire de prononciation standard, comportant peu de variantes : le nombre de phonème et leur durée moyenne en millisecondes (ms). Les résultats sont donnés pour chaque point d'enquête dans le tableau 2.4.

Le débit des locuteurs est variable selon les points d'enquête et selon le type de parole (la durée des phonèmes va de 70 à 94 ms). Cependant, d'un style à l'autre, les durées moyennes pour chaque variété restent en général comparables à 10 ms près. Au total, les phonèmes sont plus longs en lecture et plus courts en conversation libre. Des moyennes identiques peuvent toutefois cacher des débits plus ou moins réguliers correspondant à une variance faible ou élevée. La variation de débit due au style de parole est illustrée par la figure 2.3, qui présente la répartition de la durée des phonèmes en lecture et en parole spontanée (guidée ou libre). Les courbes ne sont pas définies pour les durées inférieures à 30 ms, car c'est la durée minimum imposée par le système — chaque état (3 au total) correspond au moins à la durée d'un vecteur acoustique (ici 10 ms). Il y a plus de phonèmes courts (≤ 30 ms) en parole spontanée. La distribution est également plus aplatie pour ce style de parole : celui-ci présente plus de variation que la lecture en terme de durée de phonèmes. Les courbes

	texte		entretien guidé		conversation libre	
	#phon.	durée moy. (ms)	#phon.	durée moy. (ms)	#phon.	durée moy. (ms)
Brécey	16 066	81	23 631	80	15 682	75
Brunoy	14 339	84	25 409	80	26 399	76
Dijon	10 832	80	33 320	77	31 560	78
Lyon	14 647	77	21 119	79	24 857	76
Roanne	12 417	88	20 643	85	30 302	82
Treize-Vents	10 787	77	32 951	70	28 945	71
Total Standard	**79 088**	**81**	**157 073**	**78**	**157 745**	**76**
Biarritz	17 549	80	44 528	79	14 120	76
Douzens	14 753	85	48 399	79	20 510	79
Lacaune	19 570	94	21 515	89	13 251	91
Marseille	14 645	77	0	-	0	-
Rodez	10 771	81	40 260	79	25 625	79
Total Sud	**77 288**	**84**	**154 702**	**80**	**73 506**	**81**
Boersch	**17 308**	**85**	**40 953**	**78**	**0**	**-**
Gembloux	17 609	83	27 786	79	24 060	78
Liège	17 297	82	22 419	77	26 046	73
Tournai	17 297	80	26 329	80	25 863	76
Total Belgique	**52 203**	**81**	**76 534**	**79**	**75 969**	**76**
Nyon	**17 416**	**86**	**52 176**	**89**	**23 115**	**78**
Total	**243 303**	**82**	**481 438**	**80**	**330 335**	**77**

TABLEAU 2.4. Nombre de phonèmes et durée moyenne pour le corpus PFC (pauses exclues).

montrent un peu moins de phonèmes courts et plus de phonèmes longs pour les Suisses. Nous y reviendrons au chapitre 5.

Le tableau 2.5 présente les mêmes informations (nombre de phonèmes et durée moyenne) pour le corpus CTS. En moyenne, le débit est comparable à ceux calculés pour le corpus PFC à moins de 5 ms près. Il reste également comparable entre les trois variétés standard, Sud et Alsace du corpus.

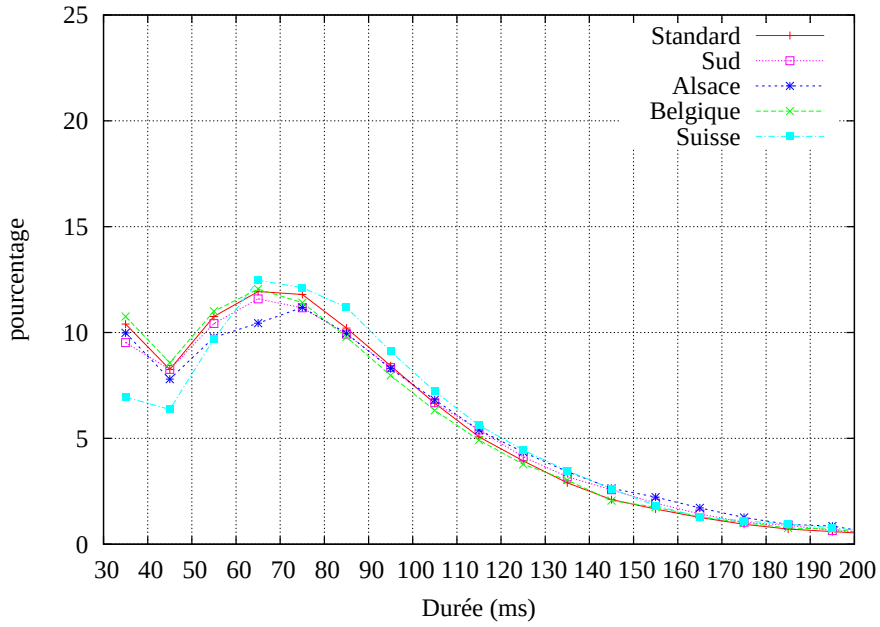
	#phonèmes	durée moyenne (ms)
Standard	819 319	80
Sud	501 577	82
Alsace	104 312	86
Total	1 425 208	81

TABLEAU 2.5. Nombre de phonèmes et durée moyenne pour le corpus CTS (pauses exclues).

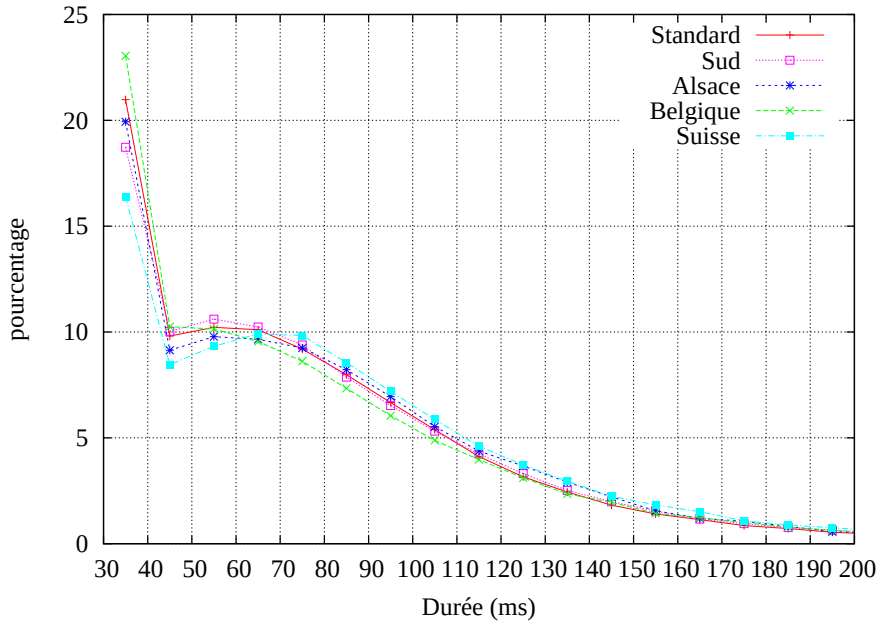
2.4 Conclusion

Ainsi, nous avons à notre disposition deux corpus d'accents régionaux, de propriétés différentes. Le corpus PFC représente plus de 40 heures de parole (provenant de 170 locuteurs) et le corpus CTS contient 65 heures de conversations téléphoniques. La structure du corpus PFC nous permet d'étudier les variations de la langue par point d'enquête, selon le style de parole et en fonction de l'âge des locuteurs. Nous étudierons également un regroupement en cinq variétés de français (standard, sud de la France, Alsace, Belgique et Suisse romande) représentées dans ce corpus, malgré des quantités de données variables. Pour trois d'entre elles (standard, sud de la France et Alsace), une validation sera permise par les données CTS.

Une partie du corpus PFC servira de base à une étude perceptive dans le chapitre suivant. Nous travaillerons ensuite sur les deux corpus dans leur intégralité : grâce à l'alignement automatique en phonèmes utilisant un dictionnaire de prononciation standard, nous extrairons du signal des caractéristiques acoustiques telles que les formants et la fréquence fondamentale. Nous étudierons enfin certaines variantes de prononciation.



(a) Lecture



(b) Parole spontanée

FIGURE 2.3. Répartition des durées des phonèmes dans le corpus PFC, pour la lecture et pour la parole spontanée.

Chapitre 3

Identification perceptive de variétés régionales de français

Dans les données audio du corpus PFC, différents accents sont représentés, soit autant de déviations par rapport à une norme. Des auditeurs natifs sont-ils capables d'identifier ces accents ? Combien d'accents une oreille d'expert ou de non-spécialiste est-elle à même de discerner ? Ces questions ne sont pas nouvelles en dialectométrie [Séguy, 1973]. Avec quel degré de granularité, par exemple, des accents méridionaux du Sud-Est et du Sud-Ouest peuvent-ils être distingués ? Qu'en est-il des accents de Belgique ? L'origine géographique des sujets interrogés pouvant influencer les réponses aux questions précédentes, il est intéressant de confronter les résultats de tests perceptifs menés en des lieux différents. Étant donné que plusieurs tranches d'âge et « styles » de parole (celui de la lecture et celui de l'entretien notamment) sont également représentés dans le corpus PFC — comme nous l'avons vu au chapitre 2 — leur influence respective sur les performances demande aussi à être quantifiée. Tels sont les objets que ce chapitre se propose d'étudier.

Les résultats obtenus dans ce chapitre pourrons nous guider par la suite : si nos auditeurs parviennent à identifier les accents que nous leur présenterons, nous en déduirons qu'il existe des différences entre ces accents et nous pourrons alors essayer de les retrouver de manière automatique dans le signal audio. Si les auditeurs ne perçoivent pas de différence entre les variétés, cela n'implique pas forcément que nous ne décèlerons aucun indice qui puisse être mesuré automatiquement, mais la tâche pourra s'avérer plus ardue.

Après une présentation du choix de la méthode employée, nous décrirons une première série d'expériences portant sur l'identification perceptive d'accents de France et de Suisse, et menées en région parisienne et en région marseillaise. Nous décrirons ensuite une seconde série d'expériences incluant en plus des accents précédemment cités d'autres accents de France et de Belgique, et menées en région parisienne uniquement. Nous présenterons la fusion des résultats des deux séries dans la quatrième section avant de conclure.

3.1 Méthode

Nous avons cherché une approche adaptée à notre problématique parmi celles qui peuvent être employées pour étudier la variation. Notre intérêt s'est particulièrement porté sur les travaux de [Clopper et Pisoni, 2004], qui traitent entre autres de l'identification perceptive de six variétés d'anglais américain. L'analyse des résultats a permis d'identifier trois groupes perçus par les auditeurs. À notre connaissance, les résultats d'une tâche de classification (clustering) perceptive similaire n'existent pas pour le français, alors qu'une distinction Nord/Sud semble évidente pour tout locuteur natif du français, même si elle dépend du degré d'accent des locuteurs. La vaste base de données issue du projet PFC permet maintenant des études systématiques : nous nous sommes donc tournée vers des tests perceptifs qui consistent à faire écouter des échantillons de parole à des auditeurs.

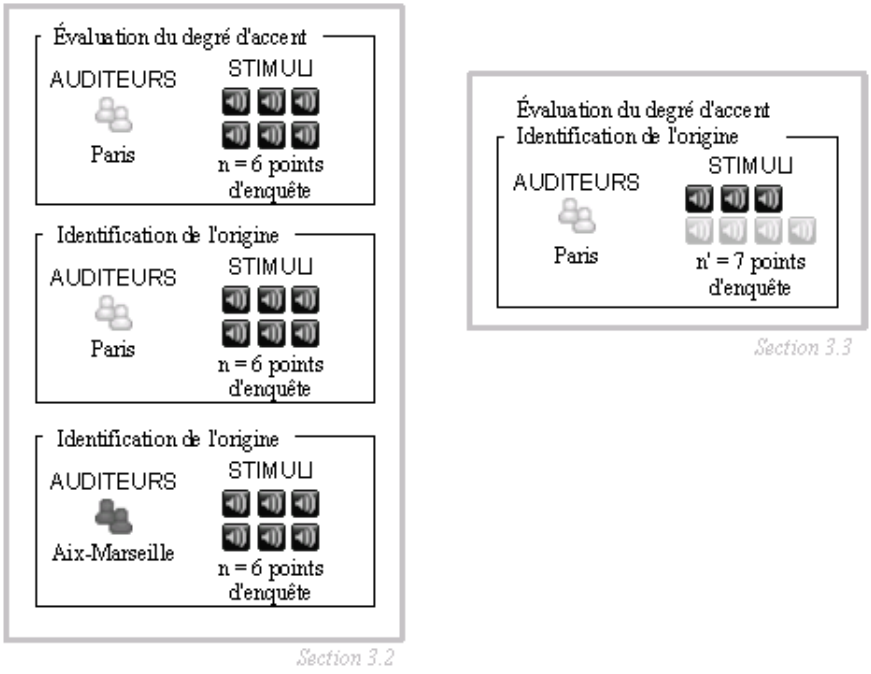


FIGURE 3.1. Mise en place des expériences perceptives.

La figure 3.1 illustre la mise en place de nos expériences. Une première série, concernant n=6 points d'enquête (Brécey, Treize-Vents, Nyon, Biarritz, Douzens et Marseille), a été menée auprès d'auditeurs de région parisienne et marseillaise. Cette série comporte une tâche d'évaluation du degré d'accent et deux tâches d'identification de l'origine, qui ont donné lieu à trois expériences distinctes. Comme chez [Clopper et Pisoni, 2004], un choix entre six possibilités, avec un niveau de hasard de 17 %, nous a semblé être un bon compromis : le nombre d'origines est assez élevé pour permettre des études, et suffisamment réduit pour ne pas allonger la durée du test et lasser les auditeurs. Nous voulions ensuite appliquer le protocole de notre première série d'expériences à des données d'Alsace (le point d'enquête de Boersch) et de Belgique (les points d'enquête de Gembloux, Liège et Tournai), tout en gardant la possibilité de pouvoir comparer nos deux séries d'expériences. Le nombre de

points d'enquête concernés passait alors de n=6 à n=10. La deuxième série concerne n'=7 points d'enquête, dont trois en commun avec la première série afin de pouvoir comparer les résultats. Ce test, comportant une tâche d'évaluation du degré d'accent et une tâche d'identification de l'origine, n'a été mené qu'en région parisienne. Nous avons pour finir fusionné les représentations graphiques des résultats des ces deux séries d'expériences.

Pour analyser les résultats des différentes expériences, nous utiliserons des mesures de rappel (équation (3.1)) et de précision (équation (3.2)). Ces deux mesures sont souvent employées dans les tâches de recherche de documents. Dans notre cas, le rappel indique le nombre de stimuli dont l'origine a été correctement identifiée par rapport au nombre total de stimuli de cette origine. Quant à la précision, elle donne le nombre de stimuli correctement attribués à une origine par rapport au nombre total de stimuli auxquels les auditeurs ont attribué cette origine. Ces deux mesures donnent des résultats compris entre 0 et 1, qui peuvent également être exprimés en pourcentages. Les chiffres qui se trouvent sur la diagonale des matrices de confusion correspondent au rappel pour chacun des points d'enquête.

$$rappel_i = \frac{\#stimuli\ correctement\ attribués\ à\ la\ région\ i}{\#stimuli\ appartenant\ à\ la\ région\ i} \tag{3.1}$$

$$précision_i = \frac{\#stimuli\ correctement\ attribués\ à\ la\ région\ i}{\#stimuli\ attribués\ à\ la\ région\ i} \tag{3.2}$$

Nous calculerons également des ANOVA (*ANalysis Of VAriance*) afin de déterminer quels facteurs ont joué un rôle dans l'identification de l'origine des échantillons.

3.2 Premières expériences : les accents de Suisse, du nord et du sud de la France

Comme mentionné dans la section précédente, notre première série d'expériences portait sur six points d'enquête PFC de Suisse, du nord et du sud de la France avec une évaluation du degré d'accent d'une quarantaine de locuteurs et des expériences d'identification de l'origine des locuteurs, menées dans la région parisienne et dans la région marseillaise. Les stimuli, le protocole et les auditeurs sont présentés dans la section 3.2.1. Les résultats de l'évaluation du degré d'accent comme des expériences d'identification (identP et identM) sont donnés en section 3.2.2 et examinés de manière plus détaillée en section 3.2.3.

3.2.1 Description des expériences

Locuteurs

Les six points d'enquête, issus du projet PFC, étaient Brécey (Normandie), Treize-Vents (Vendée), le canton de Vaud (Suisse romande), Biarritz (Pays Basque), Douzens (Languedoc) et Marseille (Provence). Dans chacun de ces points d'enquête, nous avons sélectionné six locuteurs de niveaux d'étude variés (*cf.* annexe C, p. 149), trois hommes et trois femmes,

répartis en trois tranches d'âge : 15–30 ans (moyenne : 23 ans), 30–60 ans (moyenne : 47 ans), 60 ans et plus (moyenne : 73 ans). Compte tenu de la diversité des systèmes éducatifs, du fait que les plus jeunes locuteurs n'ont pas tous terminé leurs études et que pour les plus âgés, en particulier les femmes, il était plus rare de faire de longues études, nous n'avons pas intégré le facteur scolarité dans notre travail.

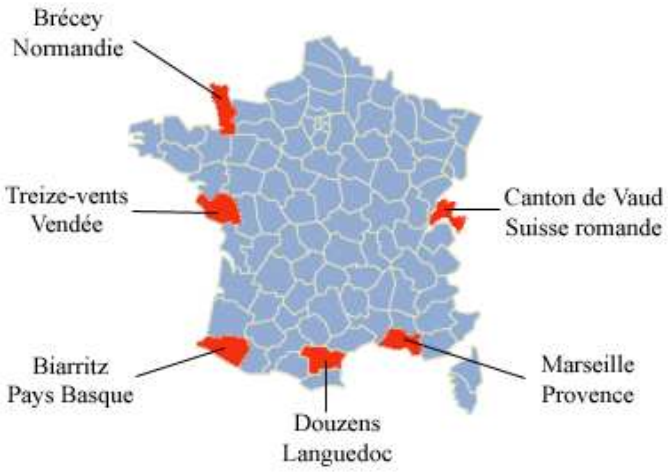


FIGURE 3.2. Zones correspondant aux six premiers points d'enquête retenus.

Comme le montre la figure 3.2, nous avons retenu trois points dans le sud de la France : Sud-Ouest, plein Sud (d'un milieu rural de vignerons) et Sud-Est (la grande ville de Marseille). Au Nord, deux points sont situés à la même latitude : Vendée et Suisse romande. Le troisième point est situé au Nord-Ouest, sur le même axe que le Pays Basque et la Vendée, en Normandie.

Stimuli

Pour chacun des locuteurs, nous avons choisi deux échantillons de parole. Le premier est une longue phrase lue (25 mots) tirée du milieu du texte PFC, identique pour tous : « La côte escarpée du mont Saint-Pierre qui mène au village connaît des barrages chaque fois que les opposants de tous les bords manifestent leur colère ». Le second est un extrait de parole spontanée, tiré d'entretiens guidés à caractère autobiographique (la transcription des énoncés spontanés sélectionnés est donnée en annexe B, p. 145) . Il a été choisi d'après les critères suivants : énoncé assertif d'une durée équivalente à celle de l'extrait lu (8-9 secondes, *cf.* tableau 3.1), absence de référence à un lieu qui biaiserait l'identification, absence d'intervention de l'interviewer et hésitations réduites au minimum. Avec en moyenne 33 mots par extrait, le débit moyen de la parole spontanée (10,6 phonèmes/seconde) est comparable à celui de la lecture (10,4 phonèmes/seconde) pour nos stimuli. Bien que la lecture présente un débit plus « uniforme » que la parole spontanée, davantage sujette à des variations internes, nous avons des quantités d'information comparables pour les stimuli de chacun des deux styles de parole.

	durée moyenne (s)	durée maximum (s)	durée minimum (s)
lecture	8,3	12,4	6,2
parole spontanée	9,4	11,0	7,4
tous les stimuli	8,9	12,4	6,2

TABLEAU 3.1. Durée des stimuli retenus.

Au total, nous avons 3 tranches d'âge $\times$ 2 sexes $\times$ 6 points d'enquête $\times$ 2 styles de parole = 72 stimuli.

Les études antérieures reposent souvent sur la seule « parole lue », même si [Clopper et Pisoni, 2004] reconnaissent eux-mêmes les problèmes que cela pose : la lecture peut ne pas refléter la façon naturelle de parler des locuteurs, dans un usage quotidien. L'utilisation de parole lue et de parole spontanée offre selon nous plusieurs avantages. D'abord, la présentation d'une même phrase lue permet de comparer toutes choses égales par ailleurs, et se prête mieux à des mesures acoustiques. En effet, la lecture exclut des indices lexicaux et syntaxiques et garantit que les différences entre locuteurs ont trait à la prononciation. La parole spontanée, quant à elle, représente un registre de langue qui reflète mieux le véritable vernaculaire des locuteurs. En utilisant les deux types de parole, il est possible de vérifier si l'un ou l'autre favorise la reconnaissance des accents, ou si les deux sont équivalents. Enfin, la présentation des deux types de stimuli dans un ordre aléatoire permet de ne pas lasser les auditeurs et de maintenir ainsi leur attention en éveil puisqu'ils n'ont pas à écouter systématiquement la même phrase.

Protocole

Une série de tests perceptifs a été menée (*cf.* colonne de gauche de la figure 3.1, p. 38), chacun soumis à vingt-cinq auditeurs décrits dans la section suivante. Les tests se déroulaient dans une chambre isolée, les auditeurs étaient munis d'un casque fermé du même modèle, le niveau d'écoute était confortable. Les stimuli, au format Wave, étaient échantillonnés à 22,05 kHz, 16 bits, mono. Leur niveau sonore avait été égalisé à l'aide du logiciel Goldwave [1], également utilisé pour la segmentation des stimuli.

Les expériences ont été réalisées à travers une interface conviviale [Vieru-Dimulescu et Boula de Mareüil, 2004], qui permet entre autres, en cliquant sur des boutons, d'entrer des informations sur la familiarité avec tel ou tel accent et de saisir les réponses. Tout d'abord, en vue d'une brève familiarisation, l'auditeur écoutait une fois la même phrase lue par un locuteur ou une locutrice (non utilisée par la suite) de chacun des six points d'enquête en question, points d'enquête qui étaient indiqués. Lors de la phase suivante, celle du test proprement dit, l'auditeur écoutait 74 stimuli, dont les deux premiers (phrases spontanées d'un locuteur du Nord et d'une locutrice du Sud) n'étaient pas comptés dans les résultats. Les 72 stimuli suivants, extraits lus ou spontanés mélangés, étaient présentés un par un dans un ordre aléatoire différent pour chaque auditeur.

1. http ://www.goldwave.com

– **Évaluation du degré d'accent.** Lors de la phase de familiarisation, un degré d'accent était donné à titre indicatif pour chaque stimulus entendu. Lors de la phase de test, l'auditeur devait attribuer un degré d'accent à l'extrait qu'il venait d'écouter. Les degrés proposés, sur une échelle à six degrés graduée de 0 à 5, étaient paraphrasés de la façon suivante :
  0 : pas d'accent ;
  1 : petit accent ;
  2 : accent modéré ;
  3 : assez fort accent ;
  4 : fort accent ;
  5 : très fort accent.
– **Expériences d'identification identP et identM.** L'origine du locuteur était indiquée pour chaque extrait entendu lors de la phase de familiarisation. Lors du test, après chaque écoute, l'auditeur devait préciser d'après l'accent l'origine du locuteur parmi les six possibilités déjà mentionnées : Brécey (Normandie), Treize-Vents (Vendée), le canton de Vaud (Suisse romande), Biarritz (Pays Basque), Douzens (Languedoc) et Marseille (Provence). Aucune indication sur l'exactitude des réponses n'était donnée.

L'auditeur pouvait prendre le temps qu'il voulait pour répondre. Chaque stimulus pouvait être réécouté, mais il était impossible de revenir en arrière une fois passé à l'extrait suivant. Chacune des trois expériences durait une vingtaine de minutes.

Auditeurs

Chaque test perceptif a été soumis à vingt-cinq auditeurs de langue maternelle française, sans troubles d'audition connus. Les sujets étaient résidents de la région parisienne, membres du LIMSI (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur) pour l'évaluation du degré d'accent et l'expérience identP, résidents de la région marseillaise et membres du LPL (Laboratoire Parole et Langage) d'Aix-en-Provence pour l'expérience identM. Nous tenons à ce propos à remercier Noël Nguyen, directeur du LPL, et son équipe pour l'accueil et l'aide qu'ils nous ont prodigués lors de la mise en place de cette expérience. Les 75 sujets se disaient quasiment tous familiers des accents de Marseille et de Suisse, quasiment tous non familiers des autres accents. S'ils pouvaient davantage être qualifiés d'experts en linguistique, les sujets de l'expérience identM ne s'estimaient pas sensiblement plus compétents pour une tâche d'identification.

Les auditeurs qui ont évalué le degré d'accent, sept femmes et dix-huit hommes, étaient âgés en moyenne de 26 ans. Ils avaient passé en moyenne 16 ans en région parisienne. Le groupe d'auditeurs de l'expérience identP comprenait neuf femmes et seize hommes, qui étaient âgés en moyenne de 32 ans, et qui n'avaient pas participé au premier test. Ils avaient passé en moyenne 21 ans en région parisienne. Enfin les auditeurs de l'expérience identM (dix-huit femmes et sept hommes) étaient âgés en moyenne de 37 ans et avaient passé en moyenne 22 ans dans la région d'Aix/Marseille – dont 11 ans à Aix même et 6 ans à Marseille même. Huit sujets avaient vécu majoritairement à Marseille, huit sujets n'y avaient jamais vécu, mais avaient longtemps vécu à Aix.

3.2.2 Résultats

Évaluation du degré d'accent

Le tableau 3.2 résume les réponses données par les auditeurs lors de la tâche d'évaluation du degré d'accent : les locuteurs ayant le moins d'accent sont les Vendéens et les Normands. Les Provençaux sont perçus comme ayant un accent moyen, les Suisses et les Basques comme ayant un assez fort accent, tandis que l'accent des Languedociens a été jugé comme le plus fort de tous.

	degré moyen	catégorie correspondante
Normandie	0,8	1 : petit accent
Vendée	1,1	1 : petit accent
Provence	2,0	2 : accent modéré
Suisse	2,5	3 : assez fort accent
Pays Basque	2,5	3 : assez fort accent
Languedoc	3,4	3 : assez fort accent

TABLEAU 3.2. Par ordre croissant, degré d'accent moyen attribué par les auditeurs aux locuteurs de chaque point d'enquête, et catégorie correspondante.

En moyenne globale, les stimuli reçoivent le degré 2. Plus les locuteurs sont âgés, plus leur accent a été jugé fort : les degrés moyens des trois catégories d'âge sont 1,4, 2,1 et 2,7. La différence est très peu marquée entre la lecture, pour laquelle le degré moyen est 2,1, et la parole spontanée, évaluée à 2,0 — les résultats seront analysés statistiquement à travers une mise en relation avec l'expérience qui suit dans la prochaine section. On peut noter que les seuls stimuli ayant obtenu le degré 5 (très fort accent), après calcul de la moyenne des réponses données, sont ceux des locuteurs qui roulaient les *r*.

Résultats de l'expérience identP (identification en région parisienne)

Tous stimuli confondus, les auditeurs ont obtenu 42,1 % de réponses exactes. Ce pourcentage, comme ceux qui suivent, a été calculé sur le nombre total de réponses données par ce groupe d'auditeurs. Toutes les origines géographiques ont été reconnues significativement mieux que le hasard [$p < 0,01$ d'après des tests de χ^2 à 5 degrés de liberté].

Le tableau 3.3 montre les résultats par point d'enquête — les chiffres sur la diagonale correspondent au rappel pour chaque point d'enquête. De tous, la Suisse romande est la mieux identifiée. Elle est suivie par la Normandie et la Vendée, souvent confondues entre elles. Les locuteurs originaires de ces deux points d'enquête ont été identifiés plus souvent comme normands que comme vendéens. Le taux de reconnaissance des locuteurs normands (supérieur à 50 %) tient donc d'une part à ce phénomène, et d'autre part au fait que les auditeurs hésitaient entre seulement deux réponses, une fois exclus la Suisse romande et le sud de la France — à mettre en relation avec le taux élevé d'identification des locuteurs vendéens

comme normands. La précision est élevée pour la Suisse (88 %), seule représentante d'un accent de l'Est. Elle n'est que de 39 % pour la Normandie et de 34 % pour la Vendée : des locuteurs à l'accent peu marqué ont pu être reconnus comme Normands ou Vendéens. Les taux d'identification de la Vendée et de la Normandie sont par conséquent à prendre avec prudence. La Suisse romande, au substrat francoprovençal mais séparée des régions françaises par une frontière étatique, n'a été que marginalement confondue avec les variétés méridionales (dans seulement 3,7 % des cas). Par ailleurs, toujours en ce qui concerne la Suisse romande, la similarité des réponses pour les énoncés lus ou spontanés pose des questions sur le rôle de la prosodie dans l'identification de l'accent suisse. La prosodie est en effet plutôt normée en lecture par rapport à la parole spontanée ; et dans les deux cas le débit de parole n'est pas plus lent qu'ailleurs en ce qui concerne nos stimuli.

origine \ réponse	Vendée	Normandie	Suisse	Pays Basque	Languedoc	Provence	total (%)	total (#)
Vendée	37	**48**	5	2	7	1	100	300
Normandie	37	**51**	2	3	6	1	100	300
Suisse	11	15	**71**	1	2	1	100	300
Pays Basque	7	6	1	36	**38**	12	100	300
Languedoc	6	2	0	25	29	**38**	100	300
Provence	12	8	2	20	**30**	29	100	300
total (%)	110	130	81	87	112	82	600	–
total (#)	327	387	242	262	338	244	–	1800

TABLEAU 3.3. Matrice de confusion sur l'ensemble des données pour 25 auditeurs de région parisienne (%). Les pourcentages sont donnés par rapport à $12 \times 25 = 300$ réponses — total reporté dans la colonne de droite tandis que la dernière ligne indique combien de fois telle ou telle réponse a été donnée.

Les trois points du Sud ont été confondus entre eux, de sorte que leur taux d'identification moyen est inférieur à 33 %. Vingt-deux des vingt-cinq auditeurs s'estimaient pourtant capables, avant le test, de reconnaître l'accent de Marseille parmi les six proposés. Ce décalage entre le discours (souvent stéréotypé) et les capacités effectives des individus soumis à une tâche d'identification n'est pas étonnant : il a également été observé dans des travaux antérieurs [Preston, 1993; Moreau et Thiam, 1995; Bauvois, 1996; Hintze *et al.*, 2000]. Contrairement à ce qui se passait pour le nord de la France, les auditeurs n'ont pas choisi l'une des trois réponses beaucoup plus fréquemment que les autres — quelle que soit l'origine des stimuli. Les stimuli du Pays Basque et de Provence ont été reconnus comme languedociens, et ceux du Languedoc comme provençaux. La réponse « Languedoc » a été la plus souvent donnée parmi celles du Sud (338 fois au lieu de 300) : la précision est effectivement plus faible pour ce point d'enquête que pour les autre (26 % contre 35 % pour la Provence et 41 % pour le Pays basque). Ce sont les locuteurs languedociens et ruraux qui sont le mieux identifiés comme méridionaux — à 92 % en additionant les

3 pourcentages des réponses correspondant au Sud. De façon symptomatique, leur accent est plus souvent associé à Marseille que celui des locuteurs de la cité phocéenne elle-même.

D'après [Coquillon *et al.*, 2000], des locuteurs marseillais et toulousains peuvent être identifiés sur la base de traits prosodiques. Nous n'avons pas observé de séparation nette entre nos locuteurs marseillais et languedociens. L'image qu'on se forge, au Nord, de l'accent marseillais pourrait être en réalité un stéréotype méridional qui va bien au-delà de Marseille. Le Languedoc est le point d'enquête qui a reçu le plus fort degré d'accent : ces deux phénomènes peuvent expliquer la confusion entre Languedoc et Provence.

Des analyses de variance (ANOVA) ont été conduites sur les réponses comptées comme correctes (1) ou incorrectes (0) avec le facteur aléatoire Sujet et les deux facteurs intra-Sujet Style et Âge. S'il n'y a pas d'effet du facteur Style (corrélation significativement positive des résultats $[T >> 1,64]$ [2] entre lu et spontané) ni d'interaction significative entre les deux facteurs, le facteur Âge se révèle avoir un effet majeur $[F(2,48) = 3,82\ ;\ p < 0,05]$. De fait, l'origine des locuteurs les plus vieux est mieux reconnue que celle des plus jeunes : pour les catégories de locuteurs des plus au moins âgés, 46,3 %, 42,5 % et 37,5 % de réponses correctes ont été observées. La différence est moins marquée entre la parole lue, donnant un score de 41,9 %, et la parole spontanée, donnant un score de 42,3 %. Le facteur intra-Sujet Sexe des locuteurs a également été analysé, mais il n'a pas d'effet majeur, même si les hommes sont légèrement mieux reconnus que les femmes (à 42,9 contre 41,3 %).

Résultats de l'expérience identM (identification en région marseillaise)

Le score moyen de 43,9 % de bonnes réponses sur l'ensemble des stimuli sera comparé aux 42,1 % de l'expérience identP. Les taux sont comparables entre les 8 auditeurs « les plus aixois » et les 8 auditeurs « les plus marseillais ».

Dans la matrice de confusion obtenue à Aix-en-Provence, on constate que pour chaque origine géographique hormis la Vendée la réponse majoritaire est la bonne — en gras dans la diagonale du tableau 3.4. La réponse « Vendée » est plus souvent donnée à la Normandie qu'à la Vendée elle-même (38 % vs 34 %), comme précédemment. Les deux réponses Vendée et Normandie ont été données plus souvent que le hasard, sûrement pour des locuteurs à l'accent moins marqué, comme lors de l'expérience précédente. Leur précision (respectivement 31 et 36 %) est d'ailleurs plutôt faible par rapport aux autres. La Suisse est toujours

2. Si on appelle b et c les variables « bonnes réponses » aux stimuli 1 et 2 (par exemple lu et spontané), alors R, la corrélation entre b et c (de taille n), est égale à :

$$R = \frac{\sum_{i=1}^{n}(b_i - \hat{b})(c_i - \hat{c})}{\sqrt{\sum_{i=1}^{n}(b_i - \hat{b})^2} \times \sqrt{\sum_{i=1}^{n}(c_i - \hat{c})^2}}$$

où $\hat{b}$ et $\hat{c}$ représentent la proportion de bonnes réponses sur chacune des deux phrases, et les b_i et c_i valent 1 ou 0 selon que la réponse est juste ou pas. Si les variables b et c sont indépendantes, alors la variable

$$T = \frac{R}{\sqrt{1 - R^2}} \times \sqrt{n - 2}$$

suit une loi de Student à $n - 2$ degré de liberté. Si l'on prend comme hypothèse alternative $R > 0$, l'indépendance est rejetée (*i.e.* l'indice de corrélation est significatif) dès que $T > 1,64$.

très bien reconnue (à 73 %) avec une bonne précision (84 %). On note la présence de zéros sur la ligne et la colonne correspondant à la Suisse : peu de confusions se sont produites entre ce point d'enquête et les autres. Les auditeurs provençaux ont obtenu de meilleurs performances que les parisiens pour les stimuli du sud de la France, même si des confusions subsistent. La réponse « Languedoc » est à nouveau celle qui a été le plus souvent donnée et qui a la plus faible précision parmis les points du Sud : 32 %, contre 44 % pour le Pays basque et 47 % pour la Provence, qui obtient la précision la plus élevée parmi ces trois points. Les auditeurs provençaux n'ont pas choisi la réponse « Provence » plus souvent que les autres, contrairement à la tendance à proposer plus souvent la réponse « c'est un X » pour les auditeurs du groupe X observée par [Moreau et Thiam, 1995].

origine \ réponse	Vendée	Normandie	Suisse	Pays Basque	Languedoc	Provence	total (%)	total (#)
Vendée	34	**47**	8	5	5	1	100	300
Normandie	38	**46**	3	6	6	2	100	300
Suisse	13	13	**73**	1	0	0	100	300
Pays Basque	10	5	2	**41**	32	10	100	300
Languedoc	6	5	0	28	**33**	29	100	300
Provence	10	13	1	13	27	**36**	100	300
total (%)	111	129	87	94	103	78	600	–
total (#)	331	388	262	281	308	230	–	1800

TABLEAU 3.4. Matrice de confusion sur l'ensemble des données pour 25 auditeurs de région marseillaise (%). Les pourcentages sont donnés par rapport à $12 \times 25 = 300$ réponses.

Comme lors de l'expérience identP, nous n'observons pas de grand écart entre l'identification des stimuli lus et spontanés (42,7 % vs 45,1 % $[T >> 1,64]$). L'identification des locuteurs selon la tranche d'âge à laquelle ils appartiennent est également comparable à celle de l'expérience identP : l'origine des plus âgés a été bien reconnue à 48,3 %, celle des 30-60 ans à 45,3 %, celle des plus jeunes à 38,0 %. Des ANOVA ont été menées comme ci-dessus. Le facteur Âge se révèle avoir un effet majeur $[F(2, 48) = 7,22\ ;\ p < 0,01]$, alors qu'il n'y a pas d'effet du Style ni d'interaction entre les deux. Même si les femmes sont un peu mieux reconnues que les hommes (44,6 % *vs* 43,2 %), contrairement à l'expérience identP, on n'a toujours pas d'effet du Sexe des locuteurs.

Comparaison entre les résultats des trois expériences

Le degré d'accent a-t-il une influence sur les réponses des auditeurs ? Les réponses des auditeurs sont-elles différentes selon leur origine ? Nous allons mettre en relation les résultats de nos trois expériences d'évaluation du degré d'accent et d'identification pour tâcher de répondre à ces questions.

Un degré d'accent a été attribué à chaque stimulus en calculant son degré moyen d'après les résultats de l'évaluation du degré, puis en l'arrondissant pour obtenir une des six catégories. Des analyses de variance ont été conduites sur les réponses des 50 auditeurs comptées comme justes (1) ou mauvaises (0) avec le facteur aléatoire Sujet et les deux facteurs intra-Sujet Style et Degré d'une part, Âge et Degré d'autre part. Dans le premier cas, seul le facteur Degré a un effet majeur $[F(5,245) = 18,1 \; ; \; p < 0,001]$. Dans le second cas, le facteur Degré $[F(5,245) = 31,8 \; ; \; p < 0,001]$ ainsi que l'interaction entre les deux facteurs Âge et Degré $[F(8,392) = 8,11 \; ; \; p < 0,001]$ sont significatifs. On observe exactement les mêmes tendances en considérant uniquement les réponses des 25 auditeurs de région parisienne ou des 25 de région marseillaise. L'interaction entre Âge et Degré n'est pas surprenante, même si rien n'empêche dans certaines circonstances que des jeunes aient davantage d'accent que leurs aînés, pour s'en démarquer, pour affirmer leur identité ou pour une autre raison.

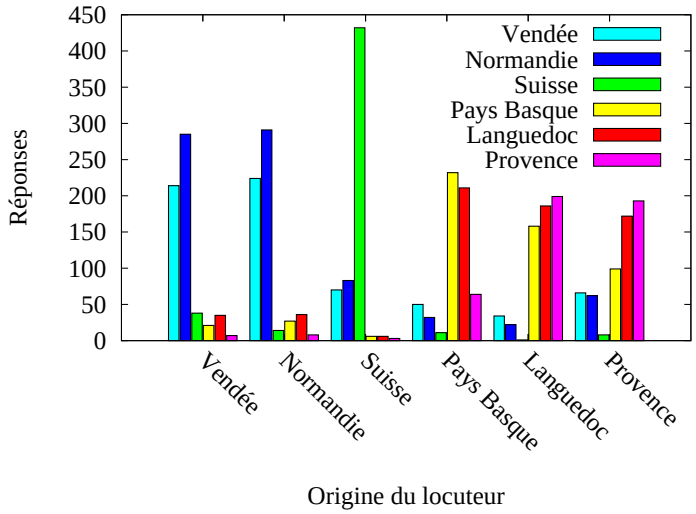


FIGURE 3.3. Résultats des expériences d'identification pour l'ensemble des 50 auditeurs (le seuil de 50 % correspond à 300 réponses).

Pour comparer les résultats des deux expériences d'identification, des ANOVA ont également été conduites avec le facteur inter-Sujet Groupe (région parisienne ou marseillaise) et le facteur intra-Sujet Âge d'une part, Degré d'autre part — les deux seuls facteurs qui jouaient un rôle d'après nos calculs. Dans les deux cas, le facteur Groupe n'a pas d'effet majeur ; seul Âge dans le premier cas $[F(2,96) = 10,4 \; ; \; p < 0,001]$ et Degré dans le second $[F(5,240) = 20,1 \; ; \; p < 0,001]$ ont un effet significatif. Il en va de même si l'on compte comme correctes des réponses confondant Normandie et Vendée, ou encore Normandie, Vendée et Suisse (*i.e.* les trois variétés non méridionales entre elles) : on n'a pas d'effet de Groupe mais un effet notamment du Degré d'accent en confondant Normandie et Vendée $[F(5,240) = 84,7 \; ; \; p < 0,001]$ et en confondant Normandie, Vendée et Suisse $[F(5,240) = 106 \; ; \; p < 0,001]$. Les auditeurs de région marseillaise étaient ainsi favorisés. Mais bien qu'ils aient légèrement mieux reconnu les accents du Sud que les auditeurs de région parisienne, leurs profils de réponses sont suffisamment semblables pour que, malgré cet arbitrage, on ne voie pas statistiquement d'influence de l'origine des auditeurs dans nos

expériences. Nous avons regroupé les réponses des 50 auditeurs dans la figure 3.3 : seul l'accent suisse est bien reconnu à plus de 50 %.

3.2.3 Analyse des résultats par clustering et scaling

Les matrices de confusion nous ont indiqué quels points d'enquêtes sont confondus entre eux par nos auditeurs. Les distances perceptives peuvent en plus être représentées sous forme graphique pour faciliter leur lecture. Des analyses statistiques ont été menées à l'aide du logiciel libre R [R Development Core Team, 2008], afin de fournir des traductions graphiques des résultats : sous forme de dendrogramme par des techniques de clustering hiérarchique, sous forme de plan à deux dimensions par échelonnement multidimensionnel (*multidimensional scaling*). Les différentes analyses nécessitent de spécifier des mesures de distance. Nous en avons utilisé deux, relativement communes : la distance euclidienne (3.3) et la distance de Manhattan (3.4), encore appelée *city block*. Leurs expressions sont données ci-dessous pour deux vecteurs X et Y de taille n.

$$\delta(X,Y) = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2} \tag{3.3}$$

$$\delta(X,Y) = \sum_{i=1}^{n} |X_i - Y_i| \tag{3.4}$$

Clustering : principe

Le clustering a pour but de regrouper des observations présentant une certaine similarité en paquets, ou *clusters*. Il s'agit d'une tâche de classification non supervisée. Le clustering hiérarchique présente la particularité de proposer des divisions en 1, 2…n classes. Ces classes peuvent être produites de deux manières à partir d'un ensemble d'observations : de bas en haut, avec un algorithme agglomératif, ou de haut en bas, avec un algorithme divisif. Ces deux techniques sont présentées plus en détail par [Manning et Schütze, 1999].

– **Algorithmes agglomératifs.** Au début, chaque observation est une classe. Ces classes sont groupées petit à petit jusqu'à former une grande classe contenant toutes les observations. À chaque itération, les deux classes les plus proches sont regroupées pour n'en former plus qu'une.
– **Algorithmes divisifs.** Au début, toutes les informations sont regroupées en une seule grande classe. Les classes sont divisées petit à petit jusqu'à ce que toutes les classes ne contiennent plus qu'un seul élément. À chaque itération, la classe qui a le plus grand « diamètre » est sélectionnée. Le diamètre d'une classe est la distance maximum entre deux de ses éléments. Pour diviser une classe, l'algorithme cherche d'abord l'observation qui est la plus éloignée du groupe — qui a en moyenne la plus grande distance avec les autres observations du groupe. Cet élément constitue une nouvelle classe. L'algorithme cherche ensuite dans les observations de l'ancienne classe celles qui sont maintenant plus

proches de la nouvelle, et les met dans la nouvelle classe. La classe d'origine est ainsi divisée en deux classes.

Ces algorithmes, contrairement par exemple à k-moyennes [MacQueen, 1967], ne dépendent pas de l'initialisation. Tous deux prennent en entrée une matrice d'observation, la matrice de confusion dans le cas présent, dont chaque ligne permet de définir un vecteur caractérisant la région correspondante. Pour chacun de ces algorithmes, il est possible de choisir une mesure de distance : euclidienne ou Manhattan. Le résultat d'un clustering hiérarchique est un dendrogramme, c'est-à-dire une représentation des données sous forme d'arbre dans lequel la distance verticale entre deux feuilles est fonction de la distance qui les sépare dans la matrice de confusion.

Scaling : principe

L'échelonnement multidimensionnel est une technique qui permet la visualisation des distances entre données, dans notre cas grâce à leur représentation dans un espace à deux dimensions. Trois algorithmes d'échelonnement multidimensionnel ont été essayés : *Classical Multidimensional scaling, Kruskal's non metric Multidimensional scaling* et *Sammon's non linear mapping*. Ces trois algorithmes prennent en entrée une matrice de dissimilitude qui peut être calculée à partir des distances entre lignes de la matrice de confusion, ceci en utilisant l'un des deux types de calcul de distance mentionnés plus haut. Cette matrice de dissimilitude, comparable aux tables de distances entre grandes villes qu'on trouve dans certains agendas, s'obtient de la façon suivante, illustrée par la figure 3.4 : la case correspondant à la ligne « Pays Basque » et à la colonne « Normandie » de la matrice de dissimilitude contient le résultat du calcul de la distance (équation (3.3) ou (3.4), p. 48) entre les lignes « Pays Basque » et « Normandie » de la matrice de confusion. Le chiffre 442 est donc obtenu par le calcul suivant en utilisant l'équation (3.3) :

$$\begin{aligned} \delta(Normandie, PaysBasque) &= \delta((111, 152, 5, 10, 19, 3), (21, 18, 3, 108, 115, 35)) \\ &= 442 \end{aligned}$$

– **Classical Multidimensional scaling.** Cet échelonnement, dit métrique, renvoie un ensemble de points représentant les données de telle sorte que les distances entre ces points soient les plus proches possibles des dissimilitudes entre les données.
– **Kruskal's non metric Multidimensional scaling et Sammon's non linear mapping.** Ces deux algorithmes sont dits non métriques : au lieu d'essayer d'approximer les dissimilitudes elles-mêmes, ils approximent une transformation monotone non linéaire de ces dissimilitudes. Ils respectent en fait seulement l'ordre relatif des dissimilitudes (monotonie). Ce sont des algorithmes itératifs, dont la configuration initiale est donnée par l'échelonnement métrique. À chaque étape, les distances entre points sont comparées aux dissimilitudes originales à l'aide d'une fonction de coût, le but étant que cette fonction prenne la plus petite valeur possible. Les deux algorithmes diffèrent par la formule de calcul du coût.

	Vendée	Normandie	Suisse	Pays Basque	Languedoc	Provence
Vendée	111	144	14	7	20	4
Normandie	**111**	**152**	**5**	**10**	**19**	**3**
Suisse	32	44	213	2	6	3
Pays Basque	**21**	**18**	**3**	**108**	**115**	**35**
Languedoc	17	6	1	75	88	113
Provence	35	23	6	60	90	86

	Vendée	**Normandie**	Suisse	Pays Basque	
Normandie	22				
Suisse	398	416			
Pays Basque	444	**442**	484		
Languedoc	478	476	518	154	
Provence	410	410	456	138	82

δ (Normandie, Pays Basque)

FIGURE 3.4. Passage de la matrice de confusion à la matrice de dissimilitude en utilisant la distance δ. Les chiffres correspondent aux nombres de réponses (donnés en pourcentages dans le tableau 3.3).

Clustering et scaling : résultats

Après cette présentation succincte des grands principes présidant aux algorithmes utilisés, nous présentons leur application à nos données sur 50 auditeurs. La figure 3.5 présente les dendrogrammes obtenus avec les deux algorithmes et les deux types de distances. Trois d'entre eux présentent le même profil : une première bipartition Nord/Sud, puis une bipartition Nord de la France/Suisse. Dans les quatre représentations graphiques de la figure 3.5, le Pays Basque se sépare des deux autres points d'enquête du Sud, mais la distance est minime. Trois variétés ressortent : la Suisse, le nord et le sud de la France.

Nous avons également appliqué les algorithmes à des sous-ensembles de données (distinguant le type de stimulus — lecture ou parole spontanée, la tranche d'âge des locuteurs et la région d'origine des auditeurs — parisienne ou marseillaise). Dans presque toutes les configurations, le clustering donne une bipartition Nord/Sud, quel que soit le sous-ensemble avec l'algorithme agglomératif utilisant une distance de Manhattan. L'algorithme divisif avec une distance euclidienne isole toujours la Suisse de la France. Avec la même métrique, l'algorithme agglomératif ne l'isole que pour le sous-ensemble des extraits de parole spontanée, évalués par les 25 auditeurs de région parisienne. Avec la distance de Manhattan, seul l'algorithme divisif fournit une bipartition France/Suisse, et ce pour les sous-ensembles de locuteurs de 30-60 ans ou les jeunes locuteurs évalués par les 25 auditeurs de région marseillaise. Seulement avec l'algorithme divisif utilisant une distance de Manhattan on a une tripartition Suisse romande, nord de la France et sud de la France, et ce sur deux sous-ensembles de données : les extraits de parole spontanée ou ceux de locuteurs de 30-60 ans évalués par les 25 auditeurs de région parisienne. Dans les autres cas, le Nord est opposé au Sud.

Le scaling, pour sa part, fait apparaître trois groupes : nord de la France, sud de la France et Suisse romande. La première dimension correspond en gros à l'axe Est-Ouest, la deuxième à l'axe Nord-Sud. Dans la figure 3.6, les axes ont été inversés : axe vertical orienté vers le bas pour placer le Nord en haut, axe horizontal orienté vers la gauche pour placer l'Ouest à

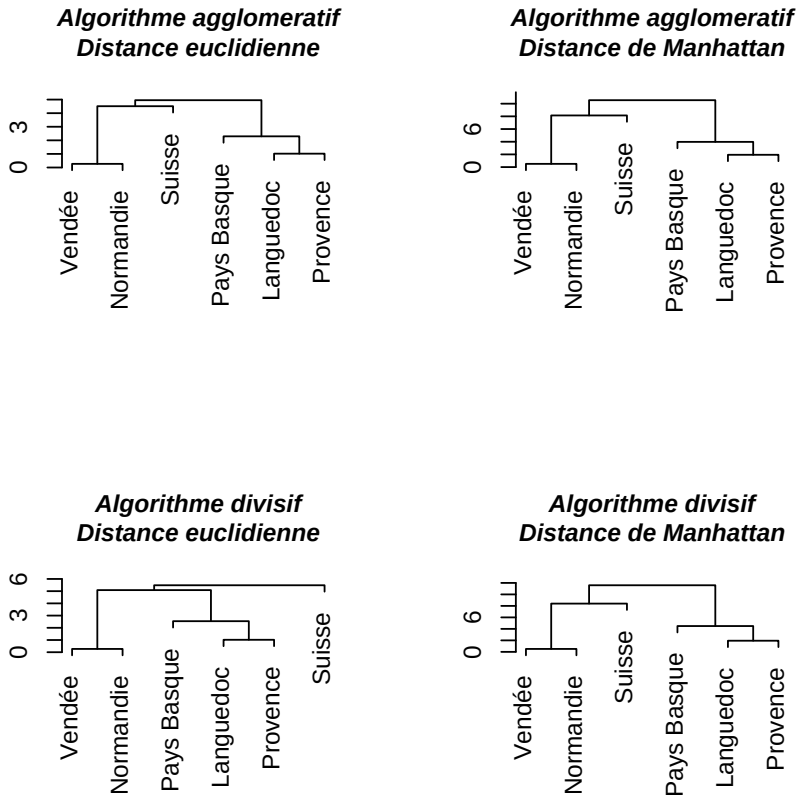


FIGURE 3.5. Dendrogrammes issus du clustering hiérarchique agglomératif ou divisif, utilisant une distance euclidienne ou de Manhattan (à partir des réponses des 50 auditeurs).

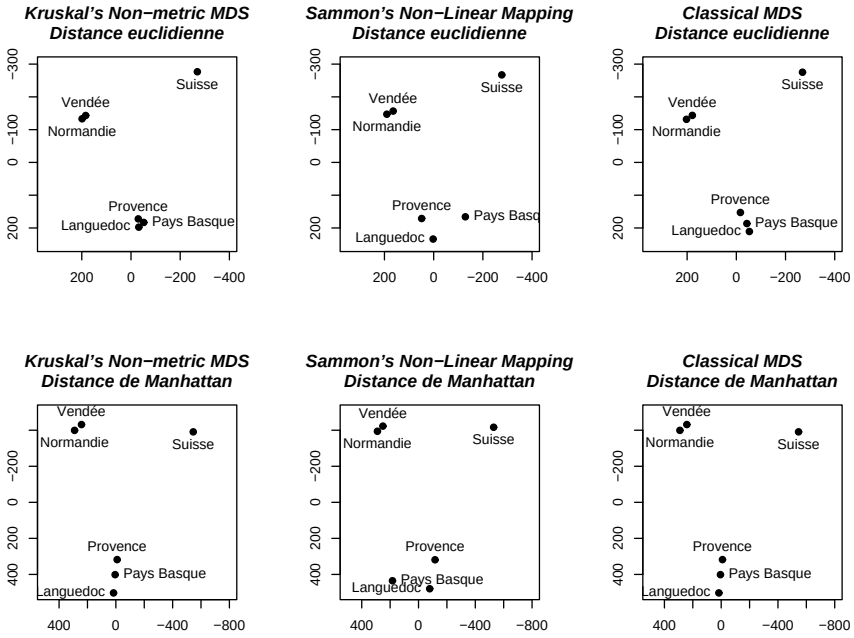


FIGURE 3.6. Résultat de l'échelonnement multidimensionnel avec les trois algorithmes et les deux distances (à partir des réponses des 50 auditeurs).

gauche, conformément à une représentation géographique classique — ce qu'il est possible de faire car ce sont les distances entre les points qui sont importantes. Quels que soient l'algorithme et la distance utilisés, sur tout ou partie des résultats, les représentations graphiques sont comparables, et les deux dimensions rendent compte d'une bonne proportion de la variance (entre 90 % et 97 % selon les cas).

3.2.4 Bilan

Les techniques de clustering et de scaling nous ont permis d'illustrer plus clairement les trois variétés de français (Suisse, nord de la France ou standard, et sud de la France) perçues par les auditeurs des deux expériences. Comme le montre le tableau 3.5, considérés séparément, chacune de ces trois variétés obtient plus de 70 % d'identification correcte. Un certain nombre de stimuli ayant un faible degré d'accent ont été identifiés comme standard : c'est ce qui explique les chiffres un peu plus élevés dans la première colonne. La précision est effectivement moins élevée pour cette variété que pour les autre (67 % contre 86 et 94 %).

Dans la section suivante, nous allons introduire de nouveaux points d'enquête. Comment les auditeurs se comporteront-ils face à des extraits sonores issus de nouveaux points d'enquête ? Ces nouveaux points d'enquête se rattacheront-ils à l'une des trois variétés que nos premiers auditeurs ont réussi à identifier ou de nouvelles variétés émergeront-elles ? La Suisse était dans ces expériences le seul point rattaché à l'Est et elle a, de ce fait, obtenu

origine \ réponse	Standard	Sud	Suisse	total (#)
Standard	**85**	11	4	1 200
Sud	15	**84**	1	1 800
Suisse	26	3	**72**	600
total (#)	1 433	1 663	504	3 600

TABLEAU 3.5. Matrice de confusion sur l'ensemble des données regroupées en 3 variétés (standard, Sud et Suisse) pour 50 auditeurs (%). Les pourcentages sont donnés par rapport au nombre de réponses de la colonne de droite.

un taux d'identification élevé. Ce ne sera peut-être pas le cas si nous ajoutons des points concurrents, comme nous l'avons fait pour les points d'enquête du Sud.

3.3 Seconde expérience incluant des accents d'Alsace et de Belgique

Une seconde expérience vient compléter les premiers tests en ajoutant de nouveaux points d'enquête, situés en Alsace et en Belgique (*cf.* colonne de droite de la figure 3.1, p. 38). Nous introduisons donc des points potentiellement concurrents pour la Suisse, ce qui met l'Est dans la même situation que le Sud dans la première série de tests.

Des auditeurs de région parisienne devaient évaluer le degré d'accent et identifier l'origine de locuteurs lors d'une seule et même expérience, qui sera présentée de la même manière que la précédente : nous décrirons d'abord les stimuli, les auditeurs et le protocole dans la section 3.3.1. Les résultats de l'expérience sont ensuite donnés en section 3.3.2, et une analyse détaillée sera fournie en section 3.3.3. Contrairement à la première fois, nous n'avons pas répliqué cette expérience auprès d'auditeurs d'origine différente.

3.3.1 Description de l'expérience

Locuteurs

Nous avons volontairement conservé certains points d'enquête utilisés dans les premiers tests, comme nous l'avions annoncé dans la section 3.1. Nous avons choisi un point d'enquête pour chacune des trois variétés identifiées dans la section précédente : Treize-Vents (Vendée) pour le nord de la France, Douzens (Languedoc) pour le Sud et Nyon (canton de Vaud) pour la Suisse romande. Au total, nous avons retenu sept points d'enquête : la figure 3.7 indique leur localisation.

FIGURE 3.7. Localisation des 7 points d'enquête retenus.

Nous n'avons conservé que quatre locuteurs par point d'enquête et ce pour plusieurs raisons. Nous avions, pour cette nouvelle expérience, un point d'enquête supplémentaire par rapport aux premiers tests (sept au lieu de six), et nous voulions éviter de trop prolonger la durée du test. De plus, les trois tranches d'âges définies précédemment n'étaient pas représentées pour le point d'enquête de Boersch : il ne nous a pas été possible d'avoir deux jeunes locuteurs pour chaque région. Or nous avons montré que l'âge des locuteurs est important pour l'identification de l'origine : nous avons donc décidé d'écarter les jeunes locuteurs (moins de 30 ans), qui étaient ceux qui, d'après nos premiers auditeurs, avaient le moins d'accent et qui étaient les moins bien reconnus. En définitive, nous avons sélectionné 4 locuteurs, deux hommes et deux femmes, répartis en deux tranches d'âge : 30-60 ans (moyenne 43 ans), et 60 ans et plus (moyenne 73 ans). Pour les points d'enquête qui avaient déjà été utilisés, nous avons gardé les mêmes locuteurs de ces deux tranches d'âge.

Stimuli

Pour chacun des locuteurs, nous avons sélectionné deux échantillons de parole de la même manière que pour les premiers tests : une phrase lue et un extrait de parole spontanée (la transcription des énoncés spontanés est donnée annexe B, p. 145). Pour les locuteurs déjà utilisés dans les premières expériences, nous avons conservé les mêmes stimuli que ci-dessus, dans la section 3.2.1. Les durées des échantillons, indiquées dans le tableau 3.6, sont comparables à celles des stimuli de la première série d'expériences (*cf.* tableau 3.1, p. 41).

Nous avons au total 2 tranches d'âge $\times$ 2 sexes $\times$ 7 points d'enquête $\times$ 2 styles de parole = 56 stimuli.

	durée moyenne (s)	durée maximum (s)	durée minimum (s)
lecture	8,8	12,4	6,4
parole spontanée	9,4	11,2	7,8
tous les stimuli	9,1	12,4	6,4

TABLEAU 3.6. Durée des stimuli retenus.

Protocole

Le test se déroulait à travers une interface accessible par Internet, contrairement aux tests précédents qui se déroulaient dans les laboratoires. De ce fait, les auditeurs ne se trouvaient plus dans l'obligation d'être en un lieu précis pour participer à l'expérience. Les auditeurs devaient écouter les échantillons avec un casque ; un champ de l'interface était prévu pour confirmer l'utilisation du casque. Les stimuli étaient au format mp3, échantillonnés à 22,05 kHz, 16 bits, mono. Leur niveau sonore avait préalablement été égalisé à l'aide du logiciel Goldwave.

Les 25 auditeurs (décrits dans la section suivante) devaient tout d'abord entrer des informations sur leur familiarité avec les accents. Lors d'une première phase de familiarisation, ils écoutaient une même phrase lue par un locuteur (non utilisé par la suite) de chacune des sept régions du test pour laquelle l'origine du locuteur était indiquée. Lors de la phase suivante, celle du test proprement dit, ils écoutaient les 56 stimuli dans un ordre aléatoire, différent pour chaque auditeur. Chaque sujet avait deux tâches à accomplir pour chaque stimuli : d'une part évaluer le degré d'accent du locuteur sur une échelle allant de 0 à 5, paraphrasée comme précédemment, et d'autre part identifier son origine parmi sept propositions : Treize-Vents (Vendée), Douzens (Languedoc), Boersch (Alsace), Nyon (Suisse romande), Gembloux (Belgique), Liège (Belgique) et Tournai (Belgique). Comme lors de la première série de tests, les auditeurs pouvaient prendre le temps qu'ils voulaient pour répondre ; ils pouvaient également réécouter les extraits sonores mais il était impossible de revenir en arrière.

Auditeurs

Le test a été soumis à 25 auditeurs de langue maternelle française, 20 hommes et 5 femmes. Ils étaient en moyenne âgés de 28 ans. Résidents de la région parisienne, ils étaient pour certains membres du LIMSI. Ils n'avaient pas de troubles de l'audition connus, et aucun d'entre eux n'avait participé à l'une de nos précédentes expériences. La moitié de nos auditeurs se disaient familiers des accents d'Alsace, du Languedoc et de Suisse, quasiment aucun d'entre eux ne se déclarait familier de l'accent de Vendée. Si plus de la moitié des auditeurs se disaient familiers de l'accent de Belgique, ils ne pensaient pas pouvoir faire la différence entre les accents de Gembloux, Liège et Tournai.

3.3.2 Résultats

Évaluation du degré d'accent

Le tableau 3.7 résume les degrés d'accent attribués en moyenne à chaque point d'enquête. Si le plus faible degré a été attribué à la Vendée, comme lors du précédent test, le degré d'accent de Tournai n'a pas été évalué comme beaucoup plus élevé. Les autres points d'enquête belges et la Suisse sont jugés comme ayant un assez fort accent, l'Alsace comme ayant un fort accent (3,7 sur 5). Comme précédemment, c'est le Languedoc qui a reçu le plus fort degré de tous (4,8 sur 5). Les degrés moyens attribués lors du précédent test (en retirant les plus jeunes locuteurs) sont indiqués entre parenthèses pour les régions concernées. Ils sont tous un peu moins élevés que les degrés attribués lors du présent test.

	degré moyen	catégorie correspondante
Treize-Vents (Vendée)	1,9 *(1,1)*	2 : accent modéré
Tournai (Belgique)	2,0	2 : accent modéré
Gembloux (Belgique)	2,7	3 : assez fort accent
Liège (Belgique)	3,0	3 : assez fort accent
Nyon (Suisse romande)	3,2 *(2,7)*	3 : assez fort accent
Boersch (Alsace)	3,7	4 : fort accent
Douzens (Languedoc)	4,8 *(4,1)*	5 : très fort accent

TABLEAU 3.7. Par ordre croissant, degré d'accent moyen attribué par les auditeurs aux locuteurs de chaque point d'enquête (entre parenthèses, degré attribué lors du premier test aux locuteurs de plus de 30 ans, s'il y a lieu), et catégorie correspondante.

En moyenne, les stimuli ont reçu le degré 3 : c'est plus que pour le précédent test, notamment en raison de la suppression des jeunes locuteurs. Les auditeurs ont peut-être également jugé plus sévèrement. Comme précédemment, le degré d'accent des locuteurs les plus âgés est plus élevé que celui des locuteurs d'âge moyen (respectivement 3,4 et 2,7). La différence est aussi très peu marquée entre les différents styles de parole : 3,0 pour la lecture et 3,1 pour la parole spontanée.

Identification de l'origine

Rappelons que lors des précédentes expériences, les trois points d'enquête du Sud étaient souvent confondus entre eux. La Suisse, seule représentante de l'Est, était très bien reconnue. Dans cette expérience, la situation est inversée : le Languedoc est la seule région représentant le Sud, et d'autres régions de l'Est peuvent générer des confusions avec la Suisse.

Tous stimuli confondus, les auditeurs ont obtenu 35,9 % de réponses correctes (pourcentage calculé sur le nombre total de réponses donné par les auditeurs). Ce taux est plus faible que

ceux obtenus lors des précédentes expériences d'identification (42,1 et 43,9 %), qui comportaient toutefois seulement six points d'enquêtes contre sept pour la présente expérience.

La matrice de confusion pour les 25 auditeurs est donnée tableau 3.8. Le Languedoc est la région la mieux identifiée avec 90 % de bonnes réponses. C'est le seul accent que les auditeurs ont reconnu quasiment à chaque fois, et qui obtient une précision élevée (83 %). On note la présence d'un 0 sur la ligne correspondante. La Vendée, la Suisse et l'Alsace viennent ensuite avec des pourcentages de réponses corrects variant entre 38 et 34 %. Leurs précisions respectives sont 33, 33 et 30 %. Les trois points d'enquête belges sont souvent confondus entre eux, ce qui explique leur faible taux d'identification. Considérés tous les trois ensembles, ils sont correctement identifiés à 49 % par les auditeurs (toutefois inférieurs aux 84 % observés pour les trois régions du Sud dans les expériences précédentes). Il est néanmoins à noter que, si la réponse la plus souvent donnée pour Liège et Gembloux est en Belgique (c'est la réponse « Gembloux »), c'est la réponse « Vendée » (représentant le français standard) qui a le plus souvent été donnée pour Tournai. Cette ville a de fait reçu le degré d'accent le plus faible parmi les points d'enquête belges.

origine \ réponse	Treize-Vents	Boersch	Nyon	Gembloux	Liège	Tournai	Douzens	total (%)	total (#)
Treize-Vents	**39**	13	10	16	11	8	5	100	200
Boersch	9	**35**	25	9	14	8	2	100	200
Nyon	4	20	**38**	14	12	11	2	100	200
Gembloux	19	14	17	**19**	16	13	4	100	200
Liège	11	18	14	**22**	19	14	3	100	200
Tournai	**29**	14	10	13	20	13	4	100	200
Douzens	6	1	2	1	0	1	**91**	100	200
total (%)	117	115	116	94	92	68	111	700	–
total (#)	231	227	228	184	180	133	217	–	1400

TABLEAU 3.8. Matrice de confusion sur l'ensemble des données pour 25 auditeurs de région parisienne (%). Les pourcentages sont donnés par rapport à $8 \times 25 = 200$ réponses.

À l'inverse des autres réponses, les trois réponses belges ont été moins souvent données que le hasard (soit 200 fois) [$p < 0,01$ d'après des tests de χ^2 à 6 degrés de liberté], notamment pour Tournai, qui n'a été choisie comme réponse que 133 fois. Ce phénomène peut en partie s'expliquer par le fait que trois réponses étaient proposées pour la Belgique, alors même que les auditeurs ne pensaient pas pouvoir faire la différence entre les trois villes. En effet, seuls quatre auditeurs ont, après le test, indiqué n'avoir pas répondu au hasard entre les trois points belges. Il semble également que les auditeurs aient globalement du mal à identifier un accent belge : une partie des réponses est répartie sur les autres points d'enquête. La réponse qui a été la plus souvent donnée est « Treize-Vents » : c'est celle qui représente le français standard. C'est également ce point d'enquête qui présente la plus faible précision

(19 %), soit un peu moins que les deux autres points de Belgique (21 %). Il faut également noter l'asymétrie de certains couples de points, notamment entre Treize-Vents et Tournai : les réponses ont été beaucoup plus nombreuses pour le représentant de la variété standard que pour Tournai, qui correspondait à une variété inconnue pour nos auditeurs.

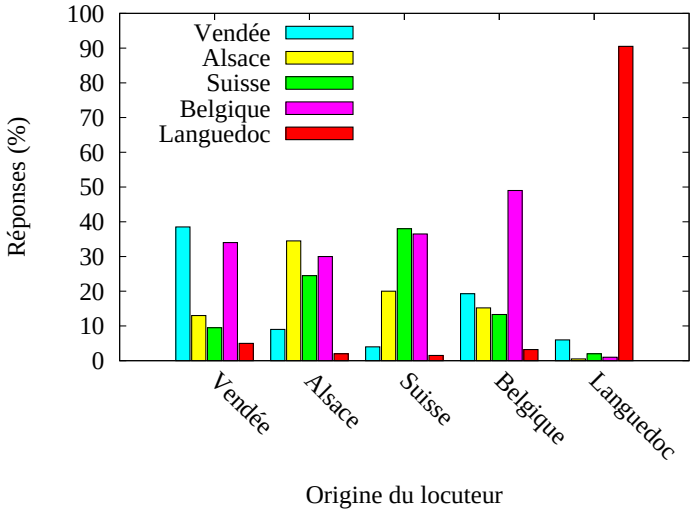


FIGURE 3.8. Résultat de l'expérience d'identification avec les points d'enquête belges regroupés pour 25 auditeurs de région parisienne.

Des ANOVA ont été conduites sur les réponses comptées comme correctes (1) ou incorrectes (0) avec le facteur aléatoire Sujet et les deux facteurs intra-Sujet Style et Âge. Le facteur Style ainsi que l'interaction des deux facteurs n'ont pas d'effet majeur. Il n'y a pas non plus ici d'effet majeur du facteur Âge contrairement à ce que nous avions observé lors des expériences précédentes. Nos locuteurs les plus âgés ont en moyenne été moins bien reconnus (à 35,6 %) que nos locuteurs d'âge moyen (36,3 %), mais l'écart entre les deux catégories est minime. Le fait d'avoir retiré une des catégories (les plus jeunes) a également éliminé une source d'information. L'écart en termes d'identification correcte est un peu plus marqué entre la lecture (37,3 %) et la parole spontanée (34,6 %), qui a ici été moins favorable à l'identification de l'origine des locuteurs.

Identification et degré d'accent

Comme pour la première série d'expériences, nous pouvons mettre en relation les résultats concernant l'identification et les degrés d'accent évalués par les auditeurs.

Un degré d'accent a, comme précédemment, été attribué à chaque stimulus en calculant d'abord son degré moyen puis en arrondissant le chiffre ainsi obtenu pour retrouver une des six catégories proposées. Avec cette méthode, aucun de nos stimuli n'a obtenu le degré 0 : nous n'avons que cinq catégories représentées. Des ANOVA ont été conduites sur les réponses comptées comme correctes (1) ou incorrectes (0) avec le facteur aléatoire Sujet et les deux facteurs intra-Sujet Style et Degré d'une part, Âge et Degré d'autre part. Dans le premier cas, nous avons trouvé un effet majeur du Style $[F(1,24) = 5,57\ ;\ p < 0,05]$ et

du Degré $[F(4, 96) = 59, 1\ ;\ p < 0, 001]$, mais aucune interaction entre les deux. Dans le second, le Degré a un effet majeur $[F(4, 96) = ;\ p < 0, 001]$ ainsi que l'interaction entre Âge et Degré $[F(4, 96) = 8, 87\ ;\ p < 0, 001]$. Ces résultats sont similaires à ceux de la première série d'expériences.

3.3.3 Analyse des résultats par clustering et scaling

Nous avons analysé les résultats du test afin d'obtenir les représentations graphiques sous forme de dendrogramme et de plan à deux dimensions. Nous avons, comme pour les résultats des expériences précédentes, utilisé le logiciel R pour appliquer des techniques de clustering hiérarchique et d'échelonnement multidimensionnel, en utilisant différents algorithmes et distances.

Quels que soient l'algorithme et la distance utilisés, les dendrogrammes donnent de façon remarquablement robuste les mêmes regroupements (figure 3.9). Comme dans l'expérience précédente, on observe tout d'abord une bipartition Nord-Est/Sud, Douzens étant toujours isolé des autres points d'enquête. La branche Nord-Est est ensuite divisée en deux groupes : Alsace et Suisse d'une part, Vendée et Belgique d'autre part. On peut remarquer que les trois points d'enquête belges sont toujours regroupés, mais que c'est Tournai, souvent identifié comme « Vendée » lors du test, qui se détache en premier.

L'échelonnement multidimensionnel oppose Douzens aux autres points selon la deuxième dimension qui correspond en gros à un axe Nord-Sud. La Vendée et la Suisse apparaissent comme les deux extrémités d'un continuum incluant l'Alsace et la Belgique. Selon l'algorithme et la métrique utilisés, ces régions sont plus ou moins regroupées ou réparties sur la première dimension, qui correspond en gros à un axe Est-Ouest en ne considérant que ses extrémités (la Belgique est placée entre Treize-Vents et Nyon). Dans la figure 3.10, l'axe des ordonnées a été inversé pour placer le Nord en haut. Dans toutes les configurations, Tournai se retrouve proche de Treize-Vents, Liège et Gembloux au milieu et enfin Boersch proche de Nyon. Les deux dimensions rendent compte de 96 à 99 % de la variance selon les cas.

3.3.4 Bilan

La première série d'expériences, qui portait sur les points d'enquête de Vendée, de Normandie, de Suisse, du Pays basque, du Languedoc et de Provence, nous a permis de mettre en évidence trois variétés de français clairement distinguées : Suisse, nord et sud de la France. La situation n'est pas la même pour notre deuxième expérience, qui inclut plus de variétés de l'Est avec les points d'enquête de Boersch, Gembloux, Liège et Tournai : si le français du Sud semble toujours bien identifié par nos auditeurs parisiens, le clustering et le scaling montrent un continuum perceptif allant des variétés standard à celles de Suisse en passant par le français de Belgique et d'Alsace. Plusieurs possiblités s'offrent à nous pour analyser les variétés faisant partie de ce continuum. Nous pourrions considérer le continuum comme une seule grande variété, mais ce serait une approximation grossière : en effet, les auditeurs des premiers tests ont été capables de différencier des points d'enquête illustrant le standard (placés à l'une des extremités) du point d'enquête suisse (situé à l'autre extremité).

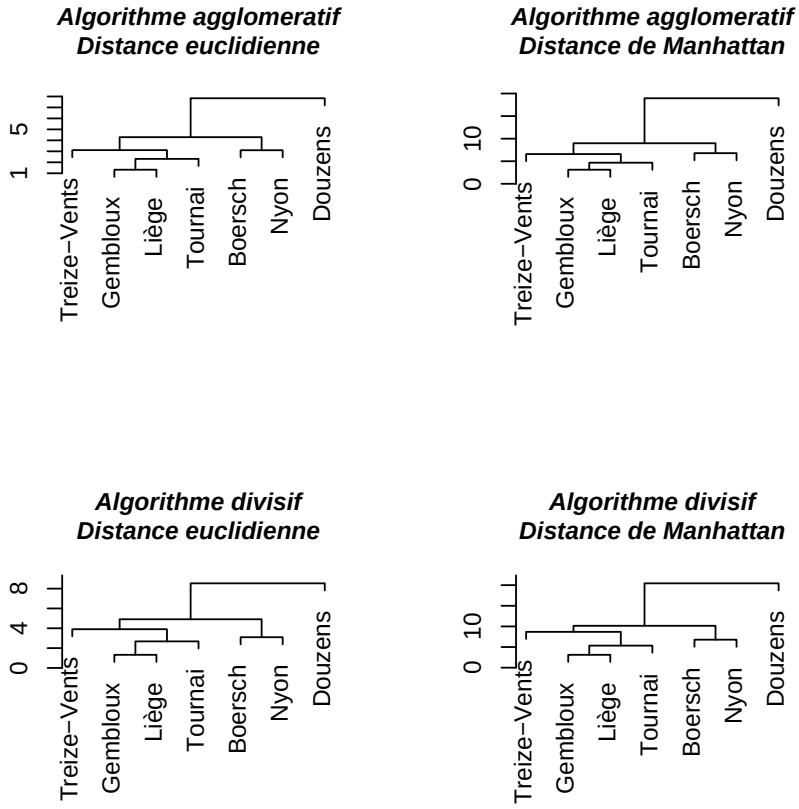


FIGURE 3.9. Dendrogrammes issus du clustering hiérarchique agglomératif ou divisif, utilisant les distances euclidienne ou de Manhattan.

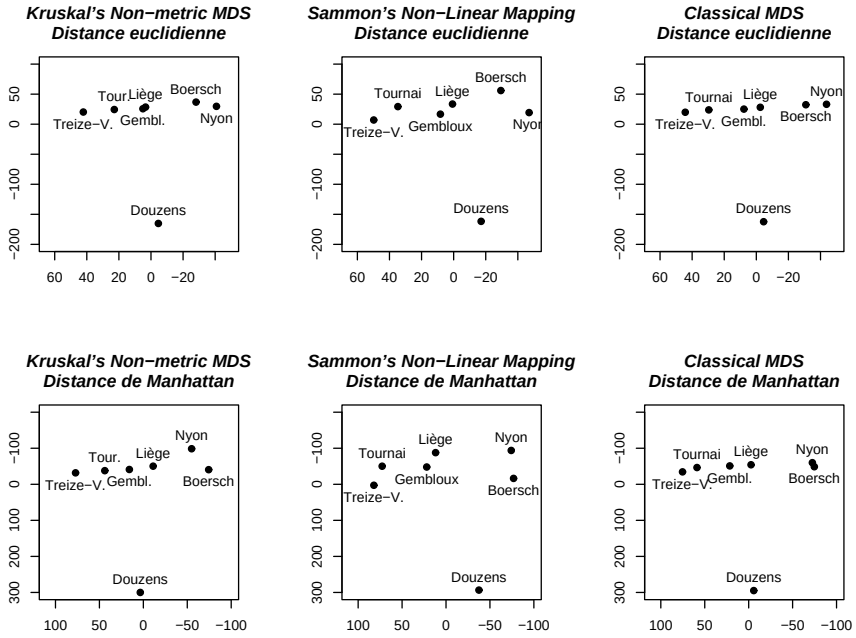


FIGURE 3.10. Résultats de l'échelonnement multidimensionnel avec 3 algorithmes et deux distances (Treize-V. = Treize-Vents, Tour. = Tournai, Gembl. = Gembloux).

Si nous voulons étudier des groupes plus réduits, nous devons placer des limites dont le positionnement reste à déterminer. Faut-il regrouper les points en fonction des frontières étatiques ? Devons-nous plutôt prendre en compte les regroupements deux à deux donnés par le clustering ?

Le tableau 3.9 donne une des matrices de confusion envisagée : les trois points d'enquête de Belgique sont inclus dans une même catégorie, tandis que les autres points restent séparés. En regroupant les données de cette manière, la réponse majoritairement donnée est pour chaque variété la bonne. Les taux d'identification sont toutefois loin d'atteindre les 70 % de l'expérience précédente pour chaque variété.

Ces résultats soulèvent plusieurs questions : les nouvelles variétés ajoutées lors de ce test auraient-elles été mieux identifiées par des auditeurs belges, alsaciens ou suisses ? Pourrions-nous trouver d'autres points d'enquête qui, de la même manière, construiraient un continuum entre le français méridional et le français standard ? Pour répondre à ces questions, de nouvelles séries de tests sont nécessaires, avec des auditeurs et des locuteurs d'autres origines. Une réplication de notre seconde expérience est en cours en Belgique.

origine \ réponse	Vendée	Alsace	Suisse	Belgique	Languedoc	total (#)
Vendée	**39**	13	10	34	5	200
Alsace	9	**35**	25	30	2	200
Suisse	4	20	**38**	37	2	200
Belgique	19	15	13	**49**	3	600
Languedoc	6	1	2	1	**91**	200
total (#)	231	227	228	497	217	1400

TABLEAU 3.9. Matrice de confusion sur l'ensemble des données regroupées en 5 variétés pour 25 auditeurs (%). Les pourcentages sont donnés par rapport au nombres de réponses de la colonne de droite.

3.4 Fusion des représentations graphiques

Nous avons obtenu une représentation graphique issue de l'échelonnement multidimensionnel pour chacune de nos expériences (*cf.* figure 3.6, p. 52 et figure 3.10, p. 61), même si les échelles sont différentes pour chaque représentation. Ces représentations permettent de visualiser les distances perçues par nos auditeurs entre les accents et elles peuvent faire émerger des regroupements. Comme il apparaît intéressant de pouvoir représenter les dix points d'enquête sur un même graphique, nous avons fusionné les représentations graphiques obtenues pour chacune des expériences afin d'obtenir une vue d'ensemble des résultats.

3.4.1 Mise en œuvre de la fusion

Nous avons transformé les coordonnées des points obtenues lors de la deuxième expérience pour faire correspondre ces nouvelles coordonnées à la représentation de la première série d'expériences. Pour ce faire, nous nous sommes servie des trois points communs aux expériences et nous avons recherché une transformation qui, prenant en entrée les coordonnées obtenues lors de la deuxième expérience, les projette dans le plan de la première série d'expérience de façon qu'ils soient les plus proches possible des trois points obtenus lors de la première série d'expérience. Tous les calculs ont été réalisés à l'aide du logiciel Matlab [3] ; nous avons cependant préféré conserver R pour les représentations graphiques par souci d'homogénéité. La mise en œuvre de la transformation est illustrée par la figure 3.11.

Nous avons posé l'équation $Y = AX + B$ (voir aussi l'équation ci-dessous), où X représente les coordonnées d'un point obtenues lors de la seconde expérience et Y ses coordonnées obtenues lors de la première expérience. Nous avons supposé qu'elle peut être résolue, ou tout du moins qu'elle admet une solution approximative. Nous avons cherché les valeurs

3. www.mathworks.fr/products/matlab/

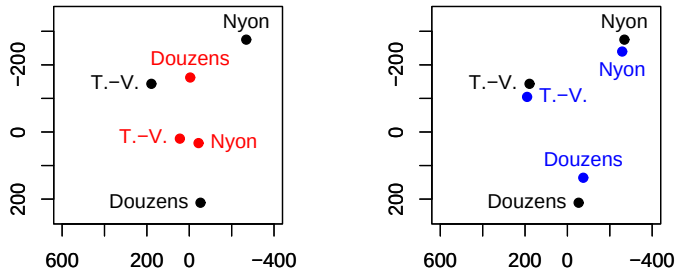


FIGURE 3.11. Mise en œuvre de la fusion — en noir, les coordonnées données par l'échelonnement multidimensionnel lors des premières expériences ; en rouge (figure de gauche), ces mêmes coordonnées obtenues lors de la deuxième expérience ; en bleu (figure de droite), les coordonnées de la deuxième expérience repositionnées pour correspondre aux premières (T.-V. = Treize-Vents).

des coefficients de A et B, de manière à ce que l'équation reste vérifiée pour les trois points que les expériences ont en commun.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Pour résoudre cette équation, nous avons tout d'abord calculé les coefficients de la matrice A d'après les valeurs de X et Y pour les trois points communs aux expériences. Nous avons obtenu de nouvelles coordonnées $X' = AX$ pour nos 3 points à projeter. Nous avons ensuite ajusté les coefficients de B d'après les valeurs des coordonnées X' et Y, pour qu'elles soient les plus semblables possible. Les nouvelles coordonnées $Z = AX + B$ des trois points sont représentées dans la figure 3.11, avec l'algorithme classique et la distance euclidienne : elles ne sont pas très éloignées des coordonnées de départ X. C'est également le cas pour les autres algorithmes et distances. Nous avons donc considéré que notre transformation était une approximation acceptable. Une fois A et B fixées, nous n'avions plus qu'à les utiliser pour appliquer la transformation aux points de la deuxième expérience que nous voulions ajouter à notre représentation graphique.

3.4.2 Résultats

Nous avons procédé de la même manière pour les 6 représentations graphiques correspondant aux trois algorithmes et aux deux types de distances pour l'échelonnement multidimensionnel : une fois la transformation définie, nous l'avons appliquée aux nouveaux points d'enquête de la deuxième expérience, à savoir Boersch, Gembloux, Liège et Tournai. Nous avons ajouté les nouveaux points ainsi obtenus aux représentations données par l'échelon-

nement multidimensionnel lors de la première série d'expériences. Le résultat est donné figure 3.12, sur laquelle nous avons homogénéisé les noms des points d'enquête : ils sont nommés d'après les villes et non d'après les régions. Les nouveaux points, indiqués en bleu, se placent sans surprise entre la Vendée et la Suisse, comme nous l'avons observé lors de la deuxième expérience.

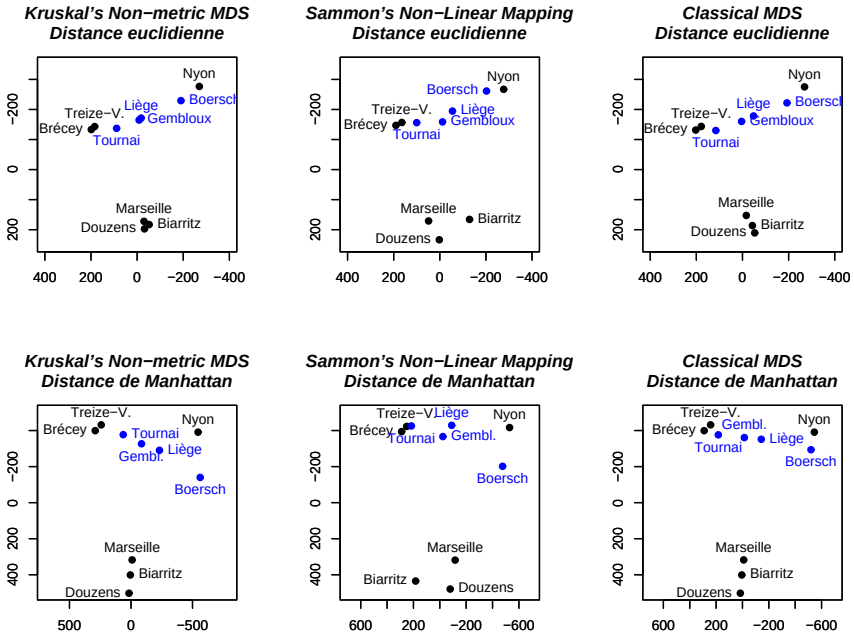


FIGURE 3.12. Résultats de la fusion des représentations graphiques pour trois algorithmes d'échelonnement multidimensionnel et les distances euclidienne ou de Manhattan. En noir, les points obtenus lors de la première série d'expériences ; en bleu, les points rajoutés.

Cette représentation permet de faire une synthèse des résultats de toutes les expériences que nous avons menées. Dans l'avenir, elle pourrait être complétée par d'autres tests perceptifs : il faudrait toutefois conserver des points d'enquête en commun pour prendre en compte l'incertitude de la méthode.

3.5 Conclusion

Dans ce chapitre, nous avons fait appel à la perception, à l'instar de [Clopper et Pisoni, 2004] ainsi qu'à des techniques d'analyse de données comme l'échelonnement multidimensionnel, à l'instar de [Heeringa, 2004]. Nous avons, dans un premier temps, dégagé trois variétés de français, Nord, Sud et Suisse, à partir de 36 locuteurs de 6 régions francophones, sans que le style (lecture formelle ou parole spontanée) ni la région d'origine (parisienne ou marseillaise) des auditeurs ne semblent affecter les résultats de manière importante. En considérant ces trois variétés, chacune d'entre elles obtient un taux d'identi-

fication correcte supérieur à 70 %. Dans un deuxième temps, à partir de 28 locuteurs de 7 régions francophones, nous avons montré que le français du nord de la France (standard) et celui de Suisse étaient placés aux deux extrémités d'un continuum incluant les accents de Belgique et d'Alsace. La distinction entre ces accents n'a pas été une tâche facile pour nos auditeurs de région parisienne. Il serait intéressant de répliquer l'expérience, notamment auprès d'auditeurs belges — le test est en cours. La fusion des représentations graphiques nous a permis d'observer ces résultats et laisse ouverte la possibilité de compléter ces études avec des locuteurs d'autres origines.

Dans la première série d'expériences, les résultats des auditeurs des régions parisienne et marseillaise convergent. Cependant, rien ne prouve que les résultats restent identiques avec d'autres locuteurs, parlant plus longtemps, ou ayant un accent plus prononcé. De même, les auditeurs auraient peut-être pu plus facilement différencier les accents de l'Est dans d'autres conditions. Mais notre but n'était pas de trouver de tels « spécimens » qui, renvoyant aux représentations, nous inscriraient dans un raisonnement circulaire. Rien n'exclut en toute rigueur qu'un accent provençal soit propre à certains locuteurs et distinct d'autres accents du sud de la France : il faut donc nuancer les conclusions. Quant à la durée des échantillons de parole, elle est dans cette expérience comparable à celle d'études antérieures [Hintze *et al.*, 2000]. Certes en l'allongeant, les chances d'identifier les accents plus finement seraient augmentées. À défaut, les résultats que nous avons obtenus avec des énoncés relativement courts (d'une dizaine de secondes) portent atteinte au mythe selon lequel un Marseillais, par exemple, qui « ne parlerait pas pareil » que d'autres Méridionaux, est facilement et immédiatement reconnu. À cet égard, malgré le biais introduit par la comparaison entre des degrés d'accent inégaux entre Marseille et un petit village du Languedoc, les fréquentes confusions entre les deux sont instructives.

Dans leur ouvrage, écrit il y a plus de 20 ans, [Carton *et al.*, 1983] considéraient une quinzaine d'accents en français. [Walter, 1982], qui reprend en grande partie les aires dialectales du territoire gallo-roman, les a encore davantage subdivisés tandis que [Martinet, 1945] regroupait les questionnaires qu'il avait recueillis en une douzaine de régions. Ceci vaut-il encore aujourd'hui ? Les accents de Provence et du Languedoc étaient distingués il y a une génération, en va-t-il encore de même à présent ?

Nos auditeurs ont aisément identifié les locuteurs du français du Sud, sans pour autant parvenir à différencier Biarritz de Marseille. Cette facilité à identifier un accent méridional se retrouvera-t-elle dans les caractéristiques acoustiques mesurées automatiquement ? Est-elle due à la composition des tests ? Arriverons-nous par la machine à différencier plus finement des accents que les humains ? Nos mesures automatiques pourraient révéler plus de différences entre les points d'enquête que les auditeurs n'en perçoivent. Les autres accents présents dans nos tests, nous l'avons vu, semblent constituer un continuum, allant du français standard à la Suisse ou à l'Alsace, en passant par la Belgique. Ceci est-il signe de difficulté pour trouver des indices permettant de différencier ces accents ? C'est ce que nous nous attacherons à étudier dans les chapitres qui suivent.

Chapitre 4

Analyse des formants

Les systèmes vocaliques sont susceptibles de présenter, selon l'accent des locuteurs, des différences que nous voudrions comparer pour nos données. La caractérisation du timbre des voyelles peut s'effectuer grâce aux mesures de formants. Pour calculer les valeurs formantiques, il est nécessaire de localiser les voyelles dans le signal de parole. Historiquement, l'étude des formants se limitait souvent à de petites quantités de données, telles que des mots isolés ou des extraits segmentés manuellement. Mais l'inconvénient que présentent des quantités de données réduites est qu'il est difficile de prendre en compte simultanément la variabilité liée au locuteur, au contexte phonémique, au style, à la variété régionale ainsi que de déterminer l'influence de chaque facteur. Si nous supposons que la segmentation par alignement phonémique permet de positionner correctement la majorité des voyelles, nous sommes en mesure de calculer les valeurs des formants dans les intervalles temporels correspondant aux voyelles pour nos deux corpus. Les différentes variétés de français ont-elles le même inventaire vocalique ? Existe-t-il des différences de timbre pour certaines voyelles ? Nous supposerons au départ que nos variétés ont toutes le même inventaire vocalique, hypothèse qui pourra ensuite être discutée au vu de nos résultats.

Aujourd'hui, plusieurs outils de traitement du signal permettant l'extraction des valeurs des formants sont disponibles. Lequel choisir ? Comment le paramétrer ? Combien de mesures prélever par voyelle ? Comment gérer les erreurs de mesure ? Quelle sera l'influence de nos choix sur la caractérisation de différents accents ? Telles sont les questions auxquelles nous tenterons de répondre dans ce chapitre, en particulier dans la section 4.1, dans laquelle est présentée une étude comparative de deux outils et de divers paramètres d'extraction sur deux variétés régionales (Nord et Sud) de français. À l'issue de cette étude et au vu des résultats qu'elle aura permis d'obtenir, nous appliquerons les mesures de formants sur d'autres variétés de français (section 4.2).

4.1 Aspects méthodologiques : étude sur les variétés nord et sud de français

L'étude présentée dans cette section se focalise sur la caractérisation de deux variétés de la même langue : le français du Nord et le français du Sud. Le but est d'étudier la variation régionale en français, mais les données d'Alsace et de Belgique n'étaient pas disponible au moment de cette étude, ce qui explique la restriction à deux variétés. L'organisation du corpus pour ce chapitre est décrite dans la section 4.1.1.

L'objectif de cette étude est également de comparer différentes méthodes d'extraction des formants. Des outils de traitement du signal — distribués gratuitement — sont aujourd'hui largement utilisés. Parmi les plus répandus, on peut citer le logiciel PRAAT [Boersma et Weenink, 2008] et la librairie SNACK Sound Toolkit [Sjölander, 2006]. Ils ont tous les deux été conçus pour permettre la création d'applications multiplateformes grâce à des langages de script. À notre connaissance, ils n'ont toutefois pas été évalués ou comparés sur de grandes bases de données [Harrison, 2004] qui ne permettent pas une annotation manuelle avec laquelle les résultats pourraient être comparés. Le choix de l'une ou l'autre technique dépend du but de la recherche.

Les techniques d'extraction de formants sont décrites dans la section 4.1.2. Les résultats sont présentés en section 4.1.3, en termes de corrélation et de distance entre les valeurs de formants. La comparaison entre PRAAT et SNACK sur nos deux corpus de parole de terrain (PFC) et de parole téléphonique (CTS) est ensuite complétée par une comparaison entre les triangles vocaliques des français du Nord et du Sud.

4.1.1 Corpus

Pour l'étude présentée dans cette section, nous avons regroupé les données que nous avions à notre disposition différemment des autres parties de notre travail. Nous allons présenter brièvement l'organisation adoptée ici.

Pour mener cette étude comparative, nous avions besoin d'une quantité relativement importante de données. Seules les variétés Nord et Sud étaient suffisamment représentées dans nos corpus PFC et CTS : nous nous sommes donc focalisée sur elles. Nous avons, pour nos deux corpus, inclus dans la variété Nord (ou standard) des données de l'Est, toujours dans le souci de disposer d'un maximum d'enregistrements.

Nous avons étudié 12 points d'enquête du corpus PFC, décrits dans le chapitre 2 : 6 dans le nord de la France (Brécey, Brunoy, Dijon, Lyon-Villeurbanne, Roanne, Treize-vents), 1 en Suisse romande (Nyon), soit au total 7 points pour la variété Nord, et 5 dans le Sud (Biarritz, Douzens, Lacaune, Marseille, Rodez). Au total, nous avons plus de 100 locuteurs, pour lesquels nous avons pris en compte la lecture et la parole spontanée disponible.

Pour cette étude, toujours dans le souci de disposer d'une base de données de grande taille, nous avons non seulement conservé les locuteurs du corpus CTS décrits dans le chapitre 2 — représentatifs du français standard et de l'Alsace —, mais nous avons également inclus des régions que nous n'utiliserons pas par ailleurs. Le total est constitué de 367 conversations

téléphoniques, regroupées en variété Nord de la Loire (Centre, Est, Nord, Ouest, Paris) et Sud (Sud-Est, Sud-Ouest). Chaque région comprend environ 70 locuteurs, dont un tiers d'hommes. Au total, le corpus représente 85 heures de parole spontanée. Il faut noter que la parole téléphonique peut poser des problèmes pour l'extraction automatique des formants et qu'elle peut nécessiter un ajustement des paramètres.

4.1.2 Méthode d'extraction des formants

Nous avons mesuré les valeurs de formants de manière automatique, à l'aide de scripts écrits pour PRAAT et SNACK. L'extraction de formants de PRAAT est fondée sur l'algorithme de Burg [Childers, 1978] ; pour SNACK, nous avons conservé l'algorithme par défaut, qui repose sur la méthode d'autocorrélation. Comme la précision de l'alignement automatique ne dépasse pas 10 ms, nous avons décidé de fonder notre analyse sur des mesures prises toutes les 10 ms.

Nous avons choisi d'extraire uniquement les trois premiers formants (F_1, F_2 et F_3), d'une part parce que notre étude pouvait s'en satisfaire, et d'autre part parce que nous étions ainsi en mesure de traiter de la même manière la parole téléphonique et non téléphonique (la fréquence maximale dans la parole téléphonique est 3300 Hz). Les réglages ont été adaptés afin de rechercher 3 formants (en dessous de 3000 Hz pour les hommes et de 3300 Hz pour les femmes avec PRAAT). La fenêtre d'analyse était réglée à 50 ms, aussi bien pour PRAAT que pour SNACK. Les autres paramètres étaient laissés à leur valeur par défaut.

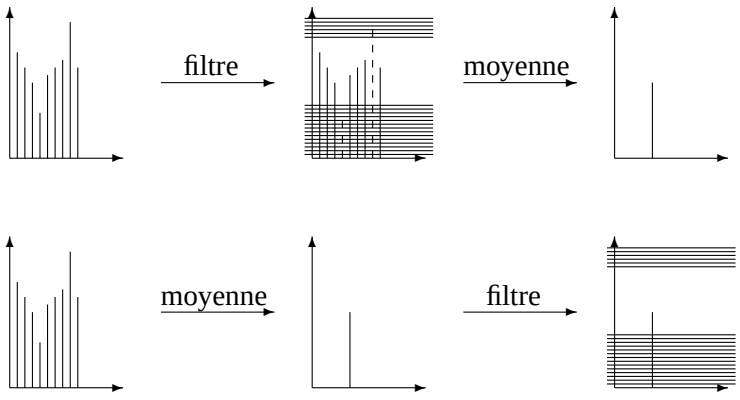


FIGURE 4.1. Application du filtre aux mesures de formants. En haut, les valeurs sont filtrées avant le calcul de la moyenne — les valeurs en pointillé sont rejetées et ne participent pas au calcul. En bas, le filtre est appliqué après le calcul de la moyenne, qui ici n'est pas rejetée.

Pour écarter les valeurs aberrantes, des filtres ont été prévus, avec des intervalles tolérants de plus ou moins 500 Hz environ par rapport à des valeurs de référence [Gendrot et Adda-Decker, 2005]. Les valeurs précises utilisées sont données en annexe D.1, p. 153. Les filtres,

adaptés à chaque voyelle, pouvaient être appliqués de deux manières différentes illustrées par la figure 4.1. Dans les deux cas, appliquer le filtre signifie éliminer les valeurs qui dépassent les limites. Les filtres peuvent être appliqués sur les valeurs brutes, avant le calcul des valeurs moyennes de formants pour chaque voyelle prononcée. Dans l'autre cas de figure, la moyenne est calculée puis le filtre est appliqué.

Des procédures de normalisation ont été proposées pour prendre en compte les caractéristiques physiologiques des locuteurs, mais elles ont leurs inconvénients [Disner, 1980; Adank, 2003], notamment le risque de masquer certaines variations intéressantes dans le cadre de notre travail. Nous avons appliqué diverses procédures de normalisation [Nearey, 1989; Syrdal et Gopal, 1986] qui n'apportaient pas grand chose de plus aux résultats. Nos corpus sont assez grands pour que des mesures en Hz puissent être présentées sans avoir recours à une normalisation intrinsèque ou extrinsèque, en passant à une échelle de type Bark.

4.1.3 Résultats

Nous avons mesuré les trois premiers formants avec PRAAT et SNACK : étant donné un fichier audio, nous avions une mesure toutes les 10 ms. Pour calculer la valeur moyenne d'un formant sur une voyelle, nous avons utilisé deux techniques : d'une part la moyenne de toutes les valeurs dans l'intervalle temporel de la voyelle, d'autre part la moyenne de trois mesures choisies à $1/4$, $1/2$ et $3/4$ de l'intervalle temporel.

Nous avons calculé les valeurs moyennes de F_1, F_2 et F_3 pour chacune des 10 voyelles orales (/a ɛ e i ɔ o u œ ø y/), par locuteur (1) d'une part, moyennées sur chacun des deux corpus (2) d'autre part. Très vite, les valeurs de F_3 se sont révélées peu fiables (beaucoup de valeurs non detectées, rejetées, ainsi qu'un écart-type élevé), c'est pourquoi nous présenterons uniquement les résultats pour les deux premiers formants dans cette section. Nous reviendrons toutefois sur les valeurs de F_3 dans la section suivante. Nous avons calculé des coefficients de corrélation (1) et des distances (2) sur chacune des voyelles considérées séparément et sur l'ensemble des voyelles.

Les coefficients de corrélation (équation (4.1)) calculés sur l'ensemble des voyelles ne sont donnés que pour les valeurs de F_1, car l'ensemble des valeurs de F_2 présente une distribution clairement non normale. On a retenu une valeur de F_1 par voyelle et par locuteur : le degré de liberté est de ce fait $ddl = 10n - 2$, n étant le nombre de locuteurs considérés. Pour les coefficients de corrélation calculés sur les voyelles isolées, $ddl = n - 2$.

$$r_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \tag{4.1}$$

Les diverses mesures de distance entre systèmes phonémiques existantes sont plutôt adaptées à des tâches de classification de phonèmes. Pour comparer deux ensembles de valeurs, par exemple, la distance de Mahalanobis [Mahalanobis, 1936] suppose qu'ils ont la même covariance. Nous avons choisi d'employer une simple distance sur les axes F_1 et F_2 considérés séparément. Les distances selon F_1 (resp. F_2) entre deux systèmes X et Y de 10 voyelles sont moyennées comme indiqué dans l'équation (4.2) (resp. (4.3)).

$$\delta_1(X,Y) = \frac{1}{10}\sum_{i=1}^{10}|F_1(X_i) - F_1(Y_i)| \tag{4.2}$$

$$\delta_2(X,Y) = \frac{1}{10}\sum_{i=1}^{10}|F_2(X_i) - F_2(Y_i)| \tag{4.3}$$

Influence du filtre et du nombre de mesures

Dans cette section, nous comparerons les mesures de formants brutes (ou sans filtre) avec les mesures filtrées, ainsi que les mesures filtrées avant ou après calcul de la moyenne. Nous présenterons les taux de rejet générés par les filtres. Nous étudierons également l'influence du nombre de mesures.

Pour quantifier l'impact du filtrage, nous avons comparé les valeurs des formants sans aucun filtre et les valeurs filtrées, que ce soit avant ou après la moyenne, sur les mesures prises toutes les 10 ms et en 3 points. Les distances $\delta(F_{brut}, F_{filtré})$ sont données dans le tableau 4.1 pour les différentes configurations. Les distances globales entre ces valeurs peuvent être assez élevées, notamment pour F_2. Elles le sont en particulier pour le F_2 extrait par SNACK sur le corpus PFC : entre 103 et 188 Hz. Les résultats voyelle par voyelle montrent des distances très élevées (plus de 100 Hz dans presque tous les cas) pour le second formant du /o/ et du /u/.

			PFC				CTS			
			hommes		femmes		hommes		femmes	
			$\delta(F_{brut}, F_{avant})$	$\delta(F_{brut}, F_{après})$	$\delta(F_{brut}, F_{avant})$	$\delta(F_{brut}, F_{après})$	$\delta(F_{brut}, F_{avant})$	$\delta(F_{brut}, F_{après})$	$\delta(F_{brut}, F_{avant})$	$\delta(F_{brut}, F_{après})$
10 ms	PRAAT	F_1	33	16	37	20	33	19	36	20
		F_2	53	32	78	49	42	24	55	35
	SNACK	F_1	10	5	8	5	5	2	3	1
		F_2	183	159	110	100	44	32	63	49
3 pts	PRAAT	F_1	31	21	35	24	33	23	35	25
		F_2	52	38	75	55	40	28	52	40
	SNACK	F_1	10	6	8	5	5	3	3	2
		F_2	188	170	109	103	43	34	60	51

TABLEAU 4.1. Distances $\delta(F_{brut}, F_{filtré})$ entre les valeurs de formants filtrées (avant ou après la moyenne) et non filtrées, pour les valeurs prises toutes les 10 ms et en 3 points, pour les corpus PFC et CTS.

Les valeurs, considérées voyelle par voyelles, sont loin d'être toujours corrélées : un tiers des coefficients de corrélation sont au dessus de 0,8 pour le corpus PFC ; la moitié pour le

corpus CTS pour les mesures prises toutes les 10 ms. Pour les mesures prises en 3 points, les proportions sont presque identiques. Les distances parfois importantes et l'absence de corrélation montrent l'utilité de l'application de filtres sur les valeurs formantiques, en particulier pour certaines voyelles comme /o/ ou /u/.

Dans la mesure où nous avons décidé de filtrer les valeurs formantiques, il est intéressant de connaître la quantité de voyelles qui ne sont pas prises en compte dans les calculs. Le taux de voyelles rejetées par les filtres (appliqués avant ou après le calcul de la moyenne) est donné dans le tableau 4.2 pour chacun des deux corpus. Ce taux est bien entendu plus élevé (7 % contre 4 % en moyenne) si les filtres sont appliqués après le calcul de la moyenne : quand ils sont appliqués avant, toutes les valeurs doivent êtres erronées pour que la voyelle soit rejetée. Le taux de rejet des valeurs moyennées sur 3 points (7 % en moyenne) est légèrement plus élevé que celui obtenu par [Gendrot et Adda-Decker, 2005] avec les mêmes critères (4 %). Bien que des critères moins sévères puissent être appliqués, la différence est marquante pour les hommes du corpus PFC, en ce qui concerne les formants mesurés avec SNACK. Les taux de rejet sont à relier aux différences entre valeurs non filtrées et valeurs filtrées : les distances entre les valeurs filtrées et non filtrées mesurées avec SNACK pour les hommes du corpus PFC sont particulièrement élevées, de même que leur taux de rejet.

			PFC		CTS	
			hommes	femmes	hommes	femmes
10 ms	PRAAT	avant	2,6	2,7	1,8	2,2
		après	6,0	6,7	4,5	5,2
	SNACK	avant	8,5	5,2	2,4	3,8
		après	14,6	9,0	4,4	6,3
3 pts	PRAAT	avant	3,6	3,7	2,3	2,7
		après	6,8	7,3	5,0	5,7
	SNACK	avant	10,7	6,5	3,0	4,6
		après	15,4	9,4	4,9	6,5

TABLEAU 4.2. Pourcentage de voyelles rejetées par les filtres appliqués avant ou après le calcul de la moyenne des mesures de formants, pour les corpus PFC et CTS.

En comparant globalement les mesures en 3 points avec les mesures prises toutes les 10 ms, nous observons un pourcentage de rejet plus élevé pour les mesures en 3 points qui pourrait être dû au fait que les deux mesures extrêmes se trouvent dans des zones de coarticulation où les mesures de formants sont moins fiables. Pour le corpus CTS, l'écart entre les deux reste cependant très faible (il varie autour de 0,5 %). Pour le corpus PFC, le taux de rejet varie un peu plus — les différences de taux de rejet peuvent atteindre 2 %. De manière générale, le corpus CTS a des taux de rejets moins élevés que le corpus PFC. Les taux sont plus faible avec PRAAT qu'avec SNACK.

Existe-t-il des différences importantes entre les mesures de formants filtrées avant et après le calcul de la moyenne ? Le calcul des distances $\delta(F_{filtré\ avant\ moyenne}, F_{filtré\ après\ moyenne})$ révèle qu'elles ne sont pas très élevées. Les corrélations sont supérieures à 0,85 et les distances moyennes sont en dessous de

30 Hz, entre mesures de formants filtrées avant ou après moyenne. Excepté pour les voyelles fermées telles que le /u/, les différences sont négligeables. Dans la suite, les filtres seront appliqués avant le calcul de la moyenne.

Les valeurs après application du filtre sont-elles différentes pour les deux types de mesures ? Les distances $\delta(F_{3\ points}, F_{10\ ms})$ ne sont dans l'ensemble pas très élevées et ne nécessitent pas d'être données en détail. Les distances moyennes calculées sur les données du corpus CTS dans les mêmes conditions d'extraction et de filtrage à partir des mesures en 3 points ou toutes les 10 ms sont inférieures à 20 Hz, aussi bien pour les hommes que pour les femmes. Pour le corpus PFC, ces mêmes distances sont toutes inférieures à 25–30 Hz. Pour les deux corpus, les distances voyelle par voyelle sont inférieures à 50 Hz sauf une (le F_2 du /o/ des femmes extrait avec PRAAT, valeurs filtrées après moyenne). Nous utiliserons dans la suite de cette étude la moyenne des mesures prises toutes les 10 ms.

Influence de l'outil

Dans la section précédente, nous avons observé que PRAAT et SNACK ne donnent pas les mêmes résultats (*cf.* tableaux 4.1 et 4.2). Dans cette section, nous cherchons à examiner les différences entre les valeurs de formants données par ces deux logiciels pour chaque voyelle. Nous utilisons des corrélations pour établir si les différences que nous observons sont dues à un décalage systématique ou aléatoire des valeurs formantiques.

Les deux dernières lignes du tableau 4.3 indiquent les distances et les corrélations entre les mesures obtenues avec PRAAT et SNACK pour l'ensemble des voyelles. Les corrélations entre les premiers formants extraits par PRAAT et SNACK sont toutes supérieures à 0,85. Les distances (somme des valeurs absolues dans les équations (4.2) et (4.3), p. 71) sont petites pour F_2 (< 50 Hz) ; elles sont plus importantes (> 50 Hz) pour F_1 sur la parole téléphonique.

Le reste du tableau 4.3 indique les différences ($F_{\text{PRAAT}} - F_{\text{SNACK}}$) et les corrélations entre les mesures obtenues avec PRAAT et SNACK pour chaque voyelle. Les corrélations entre les valeurs mesurées pour chaque voyelle par PRAAT et SNACK sont dans l'ensemble très bonnes : seulement 16 % des valeurs calculées sont en dessous de 0,7. Sur le corpus PFC, la valeur la plus basse est 0,68 et toutes les mesures semblent corrélées, même si les mesures elles-mêmes diffèrent (les distances en valeur absolue dépassent les 50 Hz pour F_1 et 100 Hz pour F_2). Les variations observées pour le corpus CTS sont plus importantes : les corrélations varient de -0,3 à 0,94 et les distances vont jusqu'à près de 100 Hz en valeur absolue. Pour une voyelle donnée, la corrélation peut être élevée et la distance importante entre les mesures obtenues avec PRAAT et SNACK (par exemple pour le F_1 du /a/ des femmes) ou inversement (par exemple pour le F_2 du /ɛ/ des femmes). Certaines mesures de F_2 ne sont que très faiblement corrélées (par exemple /a/, /œ/ et /y/ pour les femmes ; /i/, /e/ et /ɛ/ pour les hommes et les femmes). Il en résulte des triangles vocaliques de formes différentes. Pour les deux corpus, les différences calculées sur le premier formant sont toujours positives : les valeurs données par PRAAT sont toujours supérieures à celles données par SNACK, ce qui entraîne un décalage vertical des triangles vocaliques visible en comparant les figures 4.2 (p. 76) et 4.3 (p. 77).

Des t-tests ont été calculés pour chaque voyelle de la même manière que les coefficients

		PFC				CTS			
		hommes		femmes		hommes		femmes	
		cor	dist	cor	dist	cor	dist	cor	dist
a	F_1	0,91	14	0,91	42	0,94	53	0,91	93
	F_2	0,80	-77	0,93	-48	0,75	-14	0,21	-2
ɛ	F_1	0,89	17	0,87	41	0,90	54	0,87	79
	F_2	0,92	-38	0,93	-23	0,46	-21	-0,03	-18
e	F_1	0,86	19	0,80	41	0,90	40	0,80	53
	F_2	0,86	-50	0,88	-47	0,07	-25	-0,18	-37
i	F_1	0,82	28	0,75	54	0,81	51	0,64	69
	F_2	0,88	-58	0,86	-113	0,23	-53	0,07	-78
ɔ	F_1	0,88	25	0,93	43	0,89	61	0,89	86
	F_2	0,86	-36	0,97	-26	0,95	40	0,86	31
o	F_1	0,90	34	0,90	44	0,85	60	0,83	69
	F_2	0,78	1*	0,89	18*	0,89	65	0,86	58
u	F_1	0,82	45	0,78	63	0,81	69	0,78	76
	F_2	0,83	4*	0,89	32	0,81	91	0,80	83
œ	F_1	0,89	16	0,84	35	0,84	64	0,80	88
	F_2	0,86	-56	0,89	-51	0,80	27	0,50	39
ø	F_1	0,87	14	0,74	24	0,86	46	0,82	59
	F_2	0,68	-53	0,82	-54	0,89	17	0,66	10
y	F_1	0,91	22	0,82	38	0,89	58	0,89	75
	F_2	0,88	-35	0,81	-45	0,68	5	0,35	-18
tous	F_1	0,96	23	0,95	42	0,92	56	0,90	75
	F_2	–	41	–	46	–	36	–	37

TABLEAU 4.3. Corrélations et différences (en Hz) entre les mesures de PRAAT et de SNACK. Les valeurs signalées par un astérisque sont celles pour lesquelles la valeur du t-test est supérieure à 0,05 — toutes les autres peuvent donc être considérées comme statistiquement significatives.

de corrélation. Dans la majorité des cas, le t-test donne une valeur inférieure à 0,05 : les valeurs de formants obtenues locuteur par locuteur pour chaque voyelle sont différentes selon l'outil utilisé. Dans trois cas (le F_2 du /o/ pour les hommes et les femmes, et celui du /u/ pour les hommes, dans le corpus PFC), les valeurs ne sont pas statistiquement différentes ($p > 0,05$) ; pour ces voyelles, on observe des distances en Hz très petites (entre 1 et 18 Hz) entre les valeurs moyennes données par PRAAT et SNACK.

Il apparaît donc que PRAAT et SNACK présentent des différences substantielles. C'est pourquoi des comparaisons entre espaces vocaliques issus de différents outils de traitement du signal devraient être considérées avec prudence. Néanmoins, pour mettre en évidence des différences entre les systèmes vocaliques de plusieurs dialectes, variétés ou accents, PRAAT et SNACK peuvent se comporter de la même manière : cette hypothèse sera examinée dans la suite de ce travail.

Différences régionales

Une analyse plus précise des différences régionales est nécessaire. Des distances inter-corpus et inter-régions ont été calculées selon les axes F_1 et F_2. Le tableau 4.4 présente les différences Nord/Sud sur les deux corpus ainsi que les différences PFC/CTS pour les locuteurs et les locutrices du Nord et du Sud, analysés avec PRAAT et avec SNACK. Il révèle plus de différences inter-corpus qu'inter-régions, ce qui peut éventuellement s'expliquer par les particularités de la parole téléphonique.

hommes PRAAT	nord PFC	sud CTS
nord CTS	$75/55$	$6/30$
sud PFC	$16/58$	$93/61$

femmes PRAAT	nord PFC	sud CTS
nord CTS	$81/46$	$8/17$
sud PFC	$10/76$	$82/83$

hommes SNACK	nord PFC	sud CTS
nord CTS	$41/86$	$4/30$
sud PFC	$18/52$	$60/54$

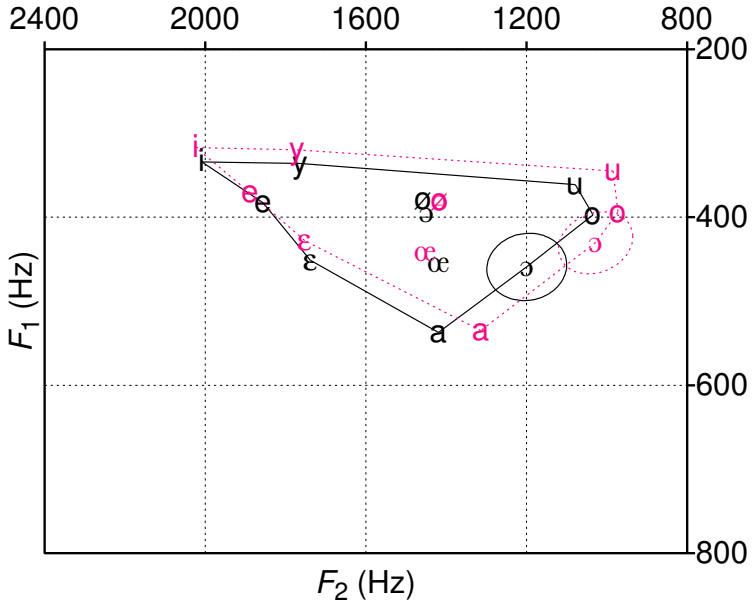
femmes SNACK	nord PFC	sud CTS
nord CTS	$45/69$	$7/23$
sud PFC	$119/73$	$55/55$

TABLEAU 4.4. Matrice des distances (selon F_1/selon F_2) obtenues pour les hommes (à gauche) et les femmes (à droite) analysées avec PRAAT (en haut) et avec SNACK (en bas) : pour chaque matrice, différences inter-corpus sur la diagonale, différences inter-régions sur l'antidiagonale.

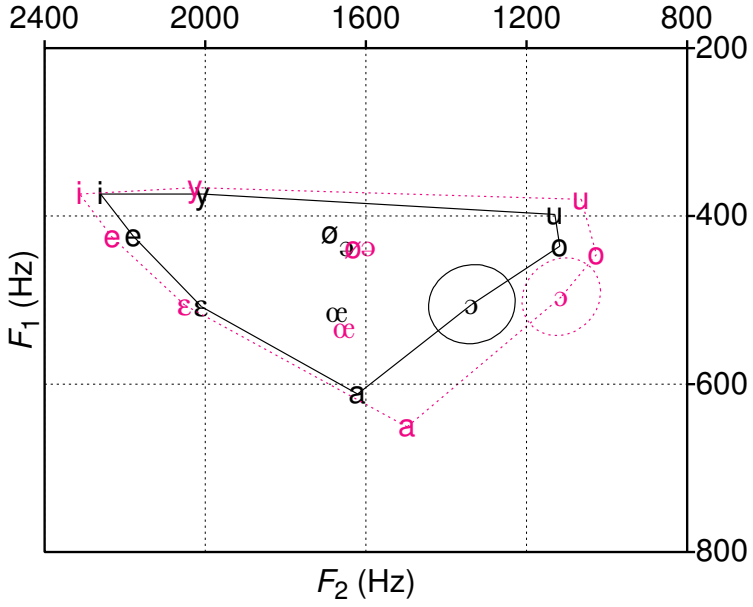
		PFC				CTS			
		hommes		femmes		hommes		femmes	
		N	S	N	S	N	S	N	S
PRAAT	F_1	450	400	500	500	500	500	600	550
	F_2	1200	1000	1350	1100	1200	1100	1350	1300
SNACK	F_1	400	400	450	450	450	450	500	500
	F_2	1250	1050	1400	1150	1150	1050	1300	1250

TABLEAU 4.5. Formants du /ɔ/ (arrondis à 50 Hz) pour les locuteurs du Nord (N) et du Sud (S).

Les triangles vocaliques ont été tracés pour les locuteurs du Nord et du Sud. Ceux issus de l'utilisation de PRAAT pour le corpus PFC sont représentés figure 4.2 ; ceux issus de l'utilisation de SNACK figure 4.3. Les triangles obtenus pour les deux variétés sont globalement comparables, avec des valeurs formantiques relativement proches pour une voyelle donnée. Quels que soient le corpus et l'outil utilisés, les hommes et les femmes montrent toutefois des différences Nord/Sud notables en ce qui concerne les voyelles postérieures, qui sont plus centralisées dans le Nord. Les locuteurs du Nord et du Sud ont des débits de parole comparables (par exemple 12,1-12,2 phonèmes/seconde pour le corpus PFC) qui ne suffisent pas à expliquer ces différences. La voyelle /a/ est en général plus haute et plus antérieure chez les locuteurs du Nord ; mais le fait le plus marquant est la réalisation du /ɔ/ qui tend à être plus fortement centralisé que les autres voyelles postérieures chez ces mêmes locuteurs. Dans

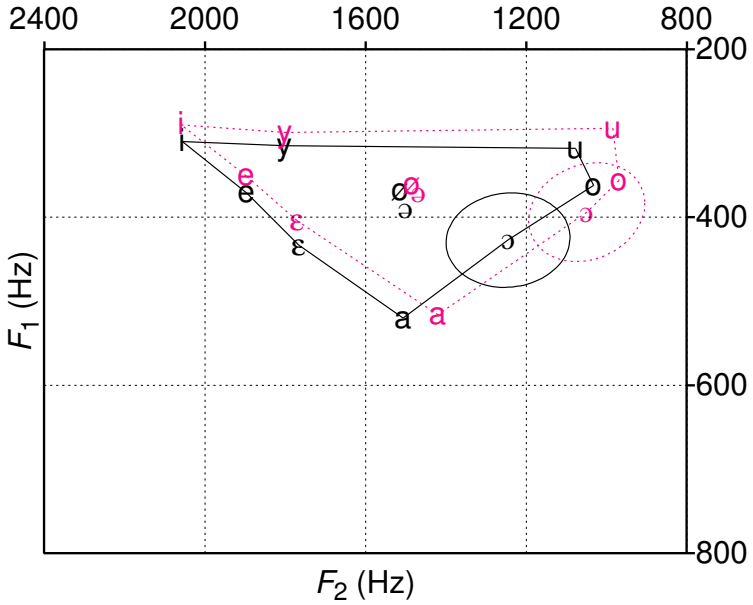


(a) Triangles vocaliques des hommes

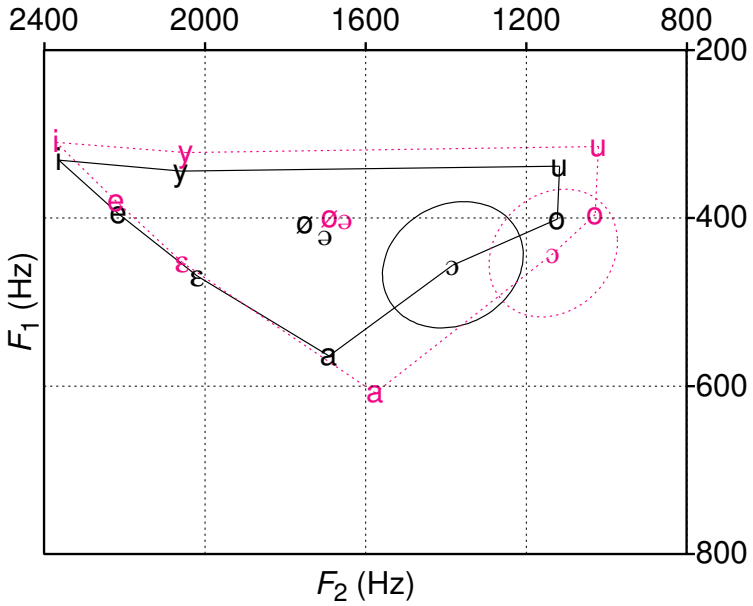


(b) Triangles vocaliques des femmes

FIGURE 4.2. Triangles vocaliques des hommes et des femmes du Nord (en ligne pleine) et du Sud (en pointillés) du corpus PFC analysé avec PRAAT. Les ellipses sont réglées à 20 % des occurences autour du /ɔ/.



(a) Triangles vocaliques des hommes



(b) Triangles vocaliques des femmes

FIGURE 4.3. Triangles vocaliques des hommes et des femmes du Nord (en ligne pleine) et du Sud (en pointillés) du corpus PFC analysé avec SNACK. Les ellipses sont réglées à 20 % des occurences autour du /ɔ/.

le Sud, le /ɔ/ est plus proche du /o/. La variation des valeurs formantiques autour de leur moyenne [Peterson et Barney, 1952] est illustrée par les ellipses de dispertion, tracées ici à 20 % autour du /ɔ/.

Comme on peut le voir dans le tableau 4.5, les différences de F_2 pour le /ɔ/ sont plus marquées sur le corpus PFC (plus de 200 Hz) que sur le corpus CTS (moins de 100 Hz). Le fait que les locuteurs de PFC soient particulièrement enracinés dans leur lieu de résidence est une explication possible.

4.1.4 Discussion

Pour les différents calculs des mesures de formants examinés, nous avons étudié l'importance du filtrage des valeurs erronées (à différents moments), fait varier le nombre de mesures prises en compte pour chaque voyelle et utilisé deux logiciels, PRAAT et SNACK, chacun mettant en œuvre des algorithmes différents.

S'il nous semble important d'écarter des mesures largement erronées en filtrant les valeurs formantiques, nous n'avons pas observé de grandes différences en faisant varier le nombre de mesures (moyenne de mesures prises toutes les 10 ms ou moyenne des mesures prises en 3 points de la voyelle) et l'application du filtre (avant ou après le calcul de la moyenne). En revanche, les valeurs des formants varient selon le logiciel (et l'algorithme) utilisé pour les mesurer ; elles varient également selon le corpus et la région d'origine du locuteur.

La conclusion à laquelle nous arrivons n'est pas surprenante : pour examiner des variations régionales dans les valeurs des formants, il convient de fixer les paramètres.

4.2 Mesures de formants sur deux corpus

L'étude menée dans la section précédente a permis de mettre en évidence des différences entre les variétés nord et sud de français. Nous allons revenir sur l'étude des formants dans nos deux corpus, en ajoutant des variétés comme la Belgique pour le corpus PFC.

4.2.1 Méthode

Pour mesurer les formants sur les nouvelles variétés, nous avons retenu le logiciel PRAAT, davantage utilisé dans la communauté des linguistes et phonéticiens, en particuliers par ceux qui travaillent sur le corpus PFC. Nous avons conservé la moyenne des mesures prises en 3 points de la voyelle et nous filtrerons les valeurs aberrantes avant de calculer la moyenne.

4.2.2 Analyse du corpus PFC

Nous avons mesuré les trois premiers formants des dix voyelles orales pour les locuteurs du français standard, du Sud, d'Alsace, de Belgique et de Suisse. Toutes les valeurs sont rapportées en annexe dans le tableau D.1, p. 154, pour les locuteurs et les locutrices du corpus

PFC. Nous avons tracé les triangles vocaliques de nos locuteurs pour la lecture (figure 4.4) et la parole spontanée (figure 4.5). Dans ces graphiques, nous avons regroupé les locuteurs d'Alsace, de Belgique et de Suisse dans une grande variété Est : la lecture s'en est trouvée simplifiée du fait de la réduction du nombre de triangles. Les triangles correspondant aux cinq variétés sont toutefois donnés en annexe D, p. 153 avec les ellipses de dispertion autour du /i/, du /a/ et du /u/.

Les mesures de formants présentent des différences selon le style de parole. Les triangles correspondant à la lecture sont plus grand que les triangles correspondant à la parole spontanée. Dans ce second style de parole, les voyelles tendent à être toutes plus centralisées. L'origine des locuteurs fait également apparaître des différences. Les triangles vocaliques sont globalement semblables pour les trois variétés que nous avons présentées, mais le comportement du /ɔ/, déjà remarqué dans la section précédente, reste différent selon les variétés : parfois antériorisé en français standard et de l'Est, il se rapproche du /o/ dans le Sud.

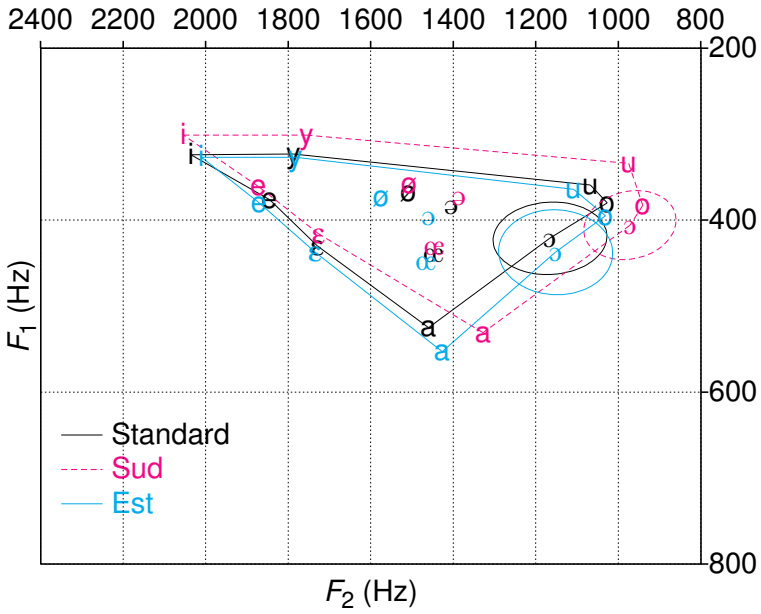
Les observations issues d'une analyse plus détaillée des formants des locuteurs d'Alsace, de Belgique et de Suisse seront à prendre avec précaution. En effet, alors que nous avions écarté les procédures de normalisation du fait d'un nombre suffisant de données pour les variétés nord et sud de français, nous ne sommes plus dans ce cas pour les variétés pour lesquelles un seul point d'enquête est disponible.

Les valeurs des deux premiers formants des locuteurs belges sont très proches de celles des locuteurs du français standard ; les triangles vocaliques de ces deux groupes de locuteurs sont très similaires. Le triangle vocalique des Suisses est un peu plus grand que celui des autres. Notamment chez les hommes, les voyelles antérieures /i/, /e/ et /ɛ/ présentent au moins l'un des deux premiers formants plus élevé par rapport au français standard. Au contraire, les voyelles postérieures /o/ et /u/ présentent un F_2 plus bas chez les locuteurs suisses. Les valeurs du premier formant des locuteurs alsaciens sont systématiquement plus élevées que celles des locuteurs des autres variétés, aussi bien pour les hommes que pour les femmes. Le deuxième formant ne semble pas être particulièrement plus élevé pour ces locuteurs. Nous n'avons pas réussi à trouver de relation entre cette différence et un indice perceptif. Ce décalage pourrait être lié aux conditions d'enregistrement, mais le manque relatif de données (par rapport au français standard, notamment) est ici regrettable : il faudrait effectuer des mesures à partir de locuteurs provenant de différents points d'enquête d'Alsace pour pouvoir faire des comparaisons. Nous ne pourrons comparer ces valeurs qu'avec celles que nous mesurerons sur le corpus CTS.

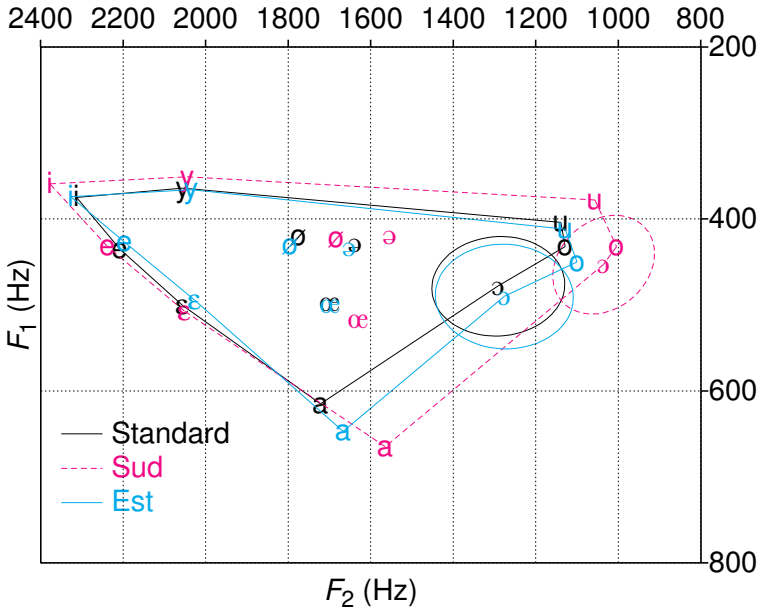
Bien que nous ayons présenté les valeurs de F_3, nous n'avons pas pu dégager de tendance claire permettant de distinguer les variétés pour ce formant.

4.2.3 Analyse du corpus CTS

Comme pour le corpus PFC, nous avons tracé les triangles pour les locuteurs et les locutrices du corpus CTS séparément (figure 4.6). Les valeurs des trois premiers formants calculées pour les voyelles des variétés standard, Sud et Alsace de français sont données en annexe dans le tableau D.2, p. 155. Les triangles obtenus pour ce corpus sont plus petits que ceux du corpus PFC. Nous avions déjà souligné, pour le corpus PFC, que les triangles correspondant

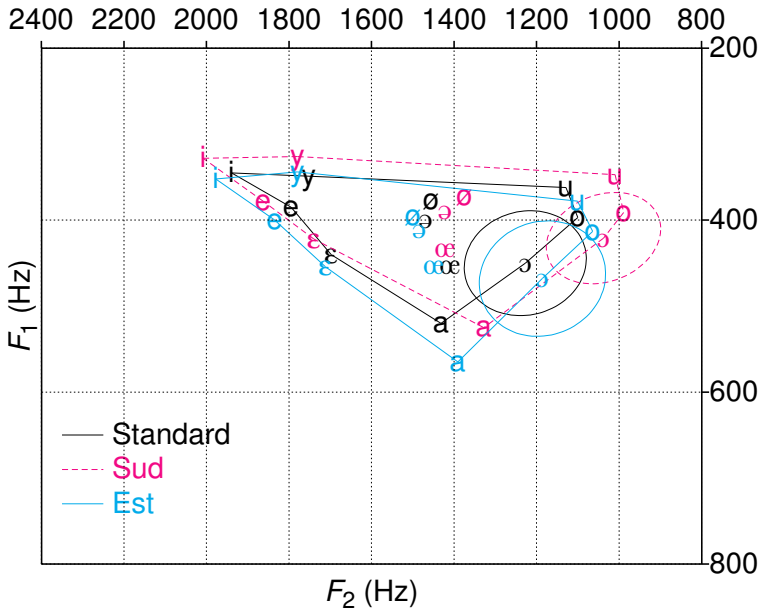


(a) Triangles vocaliques des hommes

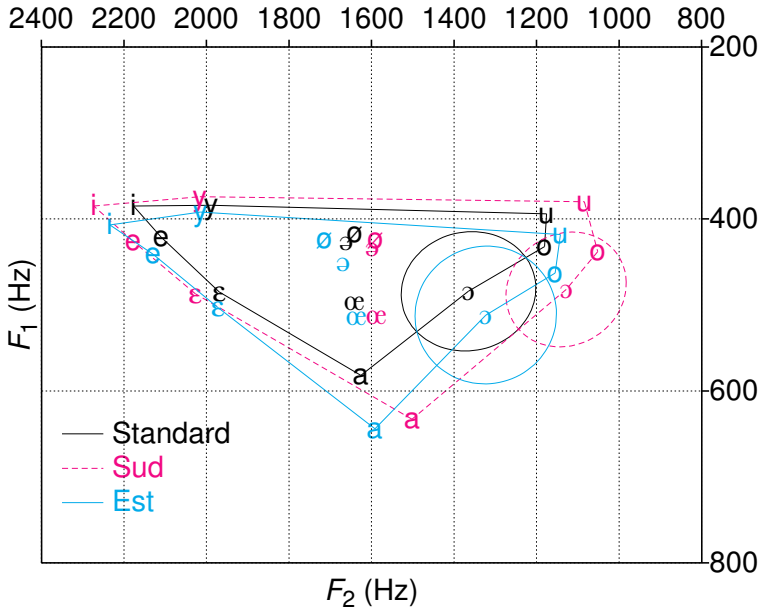


(b) Triangles vocaliques des femmes

FIGURE 4.4. Triangles vocaliques des hommes et des femmes du corpus PFC pour les locuteurs du français standard, du Sud et de l'Est (Alsace, Belgique et Suisse) pour la lecture. Les ellipses sont réglées à 20 % des occurences autour du /ɔ/.



(a) Triangles vocaliques des hommes



(b) Triangles vocaliques des femmes

FIGURE 4.5. Triangles vocaliques des hommes et des femmes du corpus PFC pour les locuteurs du français standard, du Sud et de l'Est (Alsace, Belgique et Suisse) pour la parole spontanée. Les ellipses sont réglées à 20 % des occurences autour du /ɔ/.

à la lecture sont plus grands que ceux correspondant à la parole spontanée. Le corpus CTS est constitué de conversations spontanées, ce qui peut, en partie, expliquer les dimensions des triangles. Comme nous l'avions déjà fait remarquer dans la section précédente, il sera difficile de comparer les valeurs de formants entre les corpus. Nous constatons également que le premier formant des locuteurs alsaciens est comparable à ceux des deux autres variétés présentes dans le corpus, ce qui va dans le sens de l'hypothèse d'un problème technique de ce point d'enquête dans le corpus PFC.

D'une manière générale, les valeurs formantiques des différentes variétés sont assez proches les unes des autres. Les locuteurs du français standard ont néanmoins le /ɔ/ le plus antériorisé, pour les hommes et pour les femmes. Les Alsaciens n'antériorisent pas autant le /ɔ/ ; chez les femmes, il est au niveau de celui des locutrices du sud de la France. Les Alsaciens montrent également un premier formant un peu plus élevé que les autres pour le /a/.

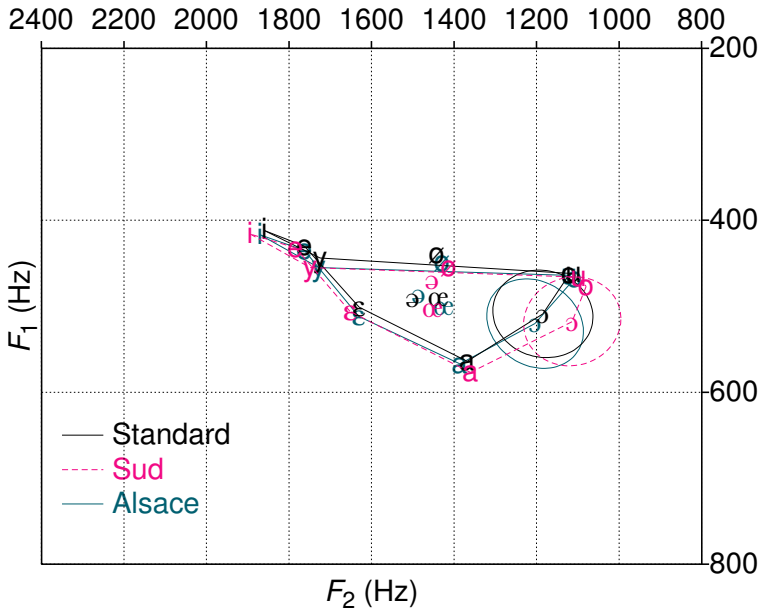
4.3 Conclusion

L'étude présentée dans la première partie de ce chapitre avait un double objectif : une comparaison entre deux outils de traitement du signal et une comparaison entre deux variétés de français (Nord et Sud). Nous avons testé deux outils d'extraction de formants sur deux grand corpus comprenant des locuteurs et des locutrices de variétés standard et non-standard de français. Après examen de l'importance du filtrage des valeurs erronées par rapport à des valeurs de référence, il est apparu que la façon d'appliquer ce filtrage ne semble pas cruciale pour nos corpus.

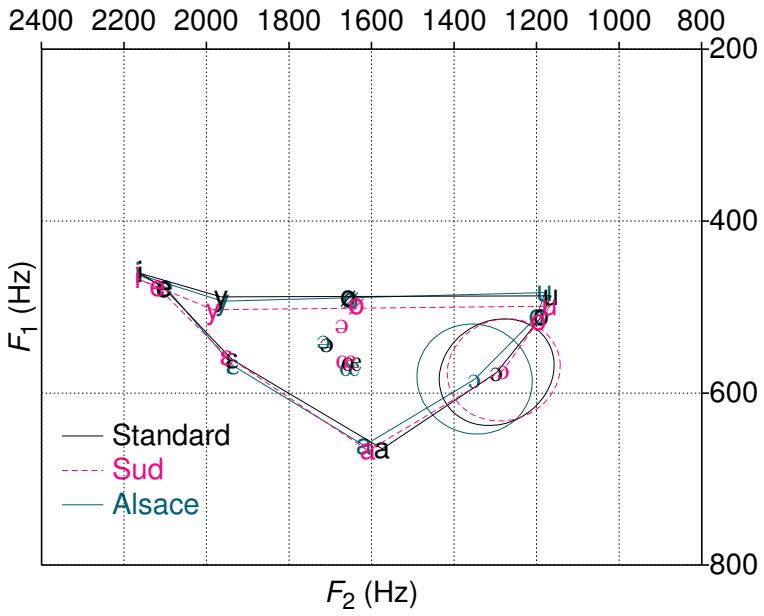
La question de l'utilisation de PRAAT ou de SNACK pour mesurer des formants sur de grands corpus s'est également posée. Les résultats sur notre corpus de parole face à face sont similaires, mais diffèrent davantage pour la parole téléphonique (qui, nous le rappelons, pose plus de problème pour l'extraction des formants), PRAAT donnant des valeurs supérieures de 50–100 Hz à celles de SNACK. Cependant, PRAAT et SNACK semblent tous deux être en mesure de faire apparaître le même type de différences entre les locuteurs du Nord et du Sud sur la base des valeurs formantiques. Il reste que l'utilisation de ces techniques nous permet de mettre en évidence l'antériorisation du /ɔ/ en français non-méridional, déjà relevée par [Martinet, 1969].

Dans un second temps, nous avons mesuré les formants sur nos corpus entiers — en incluant les points d'enquête PFC d'Alsace et de Belgique qui n'étaient pas disponibles lors de la précédente étude. Nous nous sommes heurtée au problème du manque de données : certaines mesures, notamment les valeurs de F_1 des locuteurs alsaciens, ne sont pas comparables aux autres.

Au vu des résultats, il semble qu'un paramètre soit stable pour différencier les variétés standard et Sud : le second formant du /ɔ/. Les locuteurs belges et suisses se comportent de façon comparable aux locuteurs standard en antériorisant ce phonème. Au contraire, il est réalisé d'une manière plus postériorisée et proche du /o/ chez les locuteurs du Sud. L'inventaire vocalique de ces locuteurs serait-il différent du standard ? La tendance à confondre /ɔ/ et /o/ a déjà été discutée dans la littérature [Durand, 1995].



(a) Triangles vocaliques des hommes



(b) Triangles vocaliques des femmes

FIGURE 4.6. Triangles vocaliques des hommes et des femmes du corpus CTS. Les ellipses sont réglées à 20 % des occurences autour du /ɔ/.

Les valeurs des formants ne permettent pas de caractériser tout ce qui distingue les variétés de français sur lesquelles porte ce travail. Dans les chapitres suivants, nous allons recourir à d'autres paramètres pour mettre en lumière d'autres traits de prononciation caratéristiques.

Chapitre 5

Analyse de la durée, de la fréquence fondamentale et de l'intensité

Nous nous intéressons dans ce chapitre aux paramètres de durée, de fréquence fondamentale (F_0) et d'intensité. Nous les étudierons dans des optiques différentes : d'une part pour caractériser certaines consonnes et les comparer en fonction de l'origine régionale et d'autre part pour étudier des variations prosodiques. En effet, les accents peuvent différer par des intonations ou un rythme différent. Quelques réalisations différentes du standard, localisées à certains endroits précis, peuvent suffire à donner l'impression d'un accent.

Après avoir utilisé dans le chapitre précédent les mesures de formants pour caractériser le timbre des voyelles, nous allons analyser les consonnes à l'aide de paramètres tels que la durée et la fréquence fondamentale : cette dernière peut nous donner entre autres des informations sur d'éventuels phénomènes de dévoisement. Nous étudierons ces paramètres pour les occlusives, les fricatives et le /ʁ/. Les variations des trois paramètres de durée, de fréquence fondamentale et d'intensité peuvent également porter des informations prosodiques. Ils sont par ailleurs sujets à une importante variabilité : la durée dépend du débit, la fréquence fondamentale varie selon le locuteur, notamment en fonction de son sexe, et l'intensité est tributaire non seulement du locuteur, mais aussi des conditions d'enregistrement.

Néanmoins, l'étude de la prosodie reste particulièrement intéressante pour la caractérisation des variétés de français : nous proposons des calculs prenant en compte des variations relatives et permettant de contourner les obstacles cités ci-dessus. Après une description de l'extraction des paramètres en section 5.1, nous étudierons les consonnes dans la section 5.2, puis la prosodie en section 5.3.

5.1 Extraction des paramètres

La durée des phonèmes est directement donnée par l'alignement automatique décrit précédemment (chapitre 2). Nous avons également eu recours au logiciel PRAAT [Boersma et Weenink, 2008] pour extraire F_0 et l'intensité. Comme au chapitre 4 pour l'extraction des

formants, la précision de l'alignement en phonèmes ne dépassant pas 10 ms, nous avons extrait nos deux paramètres du signal toutes les 10 ms.

La méthode d'extraction que nous avons choisie pour F_0 est fondée sur l'autocorrélation, à partir de laquelle l'algorithme détecte une périodicité acoustique [Boersma, 1993]. D'après l'auteur, cette méthode est plus robuste et moins sensible au bruit que d'autres, telles que la méthode d'autocorrélation simple ou encore les méthodes basées sur les cepstres. L'algorithme recherche une valeur de F_0 entre 75 et 600 Hz ; il peut éventuellement ne pas en trouver : dans ce cas, F_0 est non définie à l'indice temporel concerné. Lors de l'analyse du corpus CTS, nous nous heurterons au problème posé par la bande passante téléphonique (300-3300 Hz). En effet, dans cette configuration, F_0 peut être absente du signal si elle n'atteint pas 300 Hz ; dans ce cas, l'algorithme tente de la déterminer malgré tout à partir des harmoniques dans la partie du signal dont il dispose. De ce fait, nous risquons de rencontrer plus d'erreurs de détection pour ces données.

Nous avons également utilisé PRAAT pour extraire l'intensité du signal toutes les 10 ms. L'algorithme pondère la valeur de l'intensité à un moment donné par les valeurs voisines à l'aide d'une fenêtre gaussienne pour éviter de prendre en compte des variations liées aux changements de F_0. L'intensité est mesurée en dB SPL [1].

La fréquence fondamentale est variable selon les locuteurs (hommes, femmes) ; l'intensité peut, quant à elle, être affectée par les conditions d'enregistrement, la distance du locuteur par rapport au micro… Pour tenter de minimiser l'influence de ces problèmes sur nos résultats, nous nous sommes limitée d'une part à une analyse du nombre de valeurs de F_0 définies dans un intervalle temporel donné, et d'autre part à des différences de F_0 ou d'intensité entre deux voyelles consécutives pour un même locuteur.

5.2 Étude des consonnes

Dans cette section, nous avons étudié la réalisation de certaines consonnes à travers des mesures globales, à savoir la durée et le taux de voisement. Nous avons calculé ces deux paramètres pour les occlusives et fricatives sourdes /p t k f s ʃ/ ou sonores /b d g v z ʒ/ et le /ʁ/. Une durée plus importante pour les occlusives sourdes peut être l'indice d'une prononciation aspirée, et un taux de voisement particulièrement bas l'indice d'une prononciation dévoisée. Ces paramètres pourraient se montrer intéressants pour les locuteurs alsaciens : ceux que décrit [Walter, 1982] réalisent les occlusives sourdes avec aspiration, et dévoisent les plosives sonores. Certains contextes peuvent toutefois se révéler plus ou moins favorables à l'apparition de ces traits. Nous allons essayer de retrouver par nos mesures globales des traits caractérisant certaines variétés régionales.

Le taux de voisement de chaque consonne est calculé à partir des mesures de F_0 extraites du signal toutes les 10 ms. Ce taux est simplement donné par le nombre de mesures de F_0 effectivement définies par rapport au nombre maximum de mesures dans l'intervalle temporel du phonème concerné. Nous présentons les résultats de nos calculs pour le corpus PFC (section 5.2.1) et le corpus CTS (section 5.2.2).

1. Le dB SPL (*Sound Pressure Level*) est un décibel relatif à un seuil de 2×10^{-5} Pascal.

5.2.1 Analyse du corpus PFC

Le tableau 5.1 fournit le pourcentage de voisement et la durée moyenne des consonnes sourdes et sonores ainsi que du /ʁ/ pour les deux styles de parole du corpus PFC. Nous y avons ajouté les mêmes valeurs calculées pour les voyelles, qui peuvent servir de référence. Nous pouvons déjà préciser que les consonnes, regroupées en classes sonores ou sourdes par nos soins, avaient des comportements similaires examinées une à une.

		% voisement moyen				durée moyenne (ms)			
		voyelles	C sourdes	C sonores	/ʁ/	voyelles	C sourdes	C sonores	/ʁ/
lecture	Standard	88	35	86	61	81	86	72	64
	Sud	88	39	89	63	81	91	80	69
	Alsace	77	12	**48**	42	87	95	70	66
	Belgique	84	23	78	47	80	88	72	65
	Suisse	86	13	78	49	85	92	79	70
spontané	Standard	79	27	72	59	76	80	69	63
	Sud	78	28	72	56	81	83	76	67
	Alsace	70	13	**37**	46	82	83	69	60
	Belgique	62	11	41	**33**	75	81	70	63
	Suisse	68	7	46	37	88	88	77	70

TABLEAU 5.1. Taux de voisement moyen et durée moyenne (en ms) des voyelles, des consonnes sourdes, sonores et du /ʁ/ pour le corpus PFC.

Les durées ne nous permettent pas de dégager de tendance très claire selon les régions : il y a chaque fois moins de 10 ms d'écart entre les extrema atteints par les durées moyennes des consonnes. Si nous pouvons noter globalement des durées plutôt plus longues pour les consonnes sourdes des Alsaciens et des Suisses, le phénomène n'est cependant pas très marqué sur nos données : nous aurions pu nous attendre à des durées nettement plus longues pour les Alsaciens, en raison du phénomène d'aspiration des occlusives sourdes.

Les différences concernant le voisement sont plus nombreuses. Les taux de voisements sont plus élevés pour la lecture que pour la parole spontanée, dans laquelle la présence de nombreux mots outils pourrait être un facteur explicatif. Les taux de voisement varient également en fonction de l'origine des locuteurs. Dans tous les cas, ils sont comparables pour les locuteurs du français standard et du Sud. Les taux de voisement des Alsaciens, des Belges et des Suisses sont moins élevés. Dans le cas des consonnes sonores, les Alsaciens ont un taux encore plus bas que les Belges et les Suisses, notamment en lecture. Les Belges voisent moins leur /ʁ/ en parole spontanée. Toutes ces remarques sont cependant à prendre avec précaution étant données les différences observées sur les voyelles : les Alsaciens, les Belges et les Suisses présentent, selon le style de parole, un taux de voisement légèrement ou beaucoup plus bas que les autres locuteurs.

Nous pouvons également noter que nos locuteurs du français standard présentent un taux

de voisement des consonnes sourdes plus élevé et un taux de voisement des consonnes sonores moins élevé que ceux que mesure de façon similaire [Vieru-Dimulescu, 2008] sur des données plus contrôlées de français standard et avec accent étranger (18–36 % pour les consonnes sourdes, 78–94 % pour les consonnes sonores). Le taux de voisement du /ʁ/ est, lui, tout à fait comparable à l'étude citée sur les accents étrangers (59 %). Nous pouvons émettre l'hypothèse que nos locuteurs alsaciens se comporteraient comme les locuteurs allemands de [Vieru-Dimulescu, 2008] (dévoisement des sonores). En comparant les chiffres, il apparaît que le taux de voisement des consonnes sourdes et sonores pour les Alsaciens est plus bas que ceux de [Vieru-Dimulescu, 2008] pour les Allemands (soit 28–33 % et 76–93 %). Ces résultats vont dans le sens de notre hypothèse : les Alsaciens, comme les Allemands, présentent un taux de voisement des consonnes sonores moins élevé que les locuteurs de français standard. Ils montrent également que les taux de voisement de nos locuteurs alsaciens sont bas en général.

Les pourcentages de consonnes entièrement voisées et entièrement non-voisées, présentés dans le tableau 5.2, viennent compléter les taux de voisement. Les locuteurs du français standard et du Sud ont, en pourcentage, plus de consonnes entièrement voisées, toutes catégories confondues. Ils présentent également les taux les plus faibles de consonnes entièrement non-voisées. Le comportement inverse est observé pour les locuteurs d'Alsace, et dans une moindre mesure pour ceux de Belgique et de Suisse. Les Alsaciens se démarquent en lecture par encore moins de sonores voisées, ainsi que par plus de sourdes et de sonores non-voisées en lecture. Le /ʁ/ se comporte un peu différemment. En lecture, il est plutôt non-voisé pour l'Alsace, la Belgique et la Suisse ; en parole spontanée, le phénomène s'atténue pour l'Alsace tandis qu'il reste marqué pour la Belgique, suivie de la Suisse.

		% entièrement voisé			% entièrement non-voisé		
		C sourdes	C sonores	/ʁ/	C sourdes	C sonores	/ʁ/
lecture	Standard	15	64	42	27	4	19
	Sud	13	68	44	20	3	12
	Alsace	3	**30**	28	62	**27**	36
	Belgique	7	54	33	40	8	30
	Suisse	2	53	32	46	5	28
spontané	Standard	12	53	42	41	15	23
	Sud	11	54	39	38	13	21
	Alsace	7	25	30	61	37	35
	Belgique	5	25	**21**	68	37	**47**
	Suisse	3	28	22	72	31	38

TABLEAU 5.2. Pourcentage de consonnes sourdes, sonores et de /ʁ/ complètement voisées et complètement non-voisées pour le corpus PFC.

Le travail de [Hallé et Adda-Decker, 2007] nous donne des chiffres en contexte ou non d'assimilation (le trait voisé ou non-voisé de la consonne concernée dépend alors de ce même trait pour une consonne adjacente). Bien que nos pourcentages soient calculés sur le corpus entier, avec des contextes non contrôlés, nous pouvons comparer les ordres de grandeur. Les pourcentages de consonnes sourdes et sonores complètement voisées mesurés sur la variété standard sont comparables à ceux de cette étude. Il en va de même pour le pourcentage de

consonnes sonores complètement dévoisées. Au contraire, nos pourcentages de consonnes sourdes complètement non-voisées sont très élevés en comparaison des chiffres donnés par [Hallé et Adda-Decker, 2007] (20 % dans le contexte le plus favorable au dévoisement). Un problème de détection de F_0 pourrait en être à l'origine.

Nos mesures tendent à montrer que les taux de voisement varient selon les régions : deux ou trois groupes apparaissent. Les locuteurs du français standard et du Sud voisent en général plus, quels que soient les phonèmes concernés, tandis que les locuteurs des autres variétés voisent moins. L'hypothèse selon laquelle les Alsaciens voisent moins les consonnes sonores n'est pas infirmée, mais l'écart avec les autres variétés n'est pas toujours très marqué. Les Belges présentent des /ʁ/ un peu moins voisés que les autres en parole spontanée. Les différences de voisement sont parfois également observées sur les voyelles, ce qui laisse penser que la méthode pourrait avoir un biais généré par les conditions d'enregistrement plutôt que par une prononciation particulière. En effet, les trois groupes qui présentent des taux de voisement plutôt faibles sont également ceux pour lesquels nous disposons du moins de données audio. Ces tendances demanderaient à être vérifiées sur des points d'enquête supplémentaires.

5.2.2 Analyse du corpus CTS

Les pourcentages de voisement et les durées moyennes des voyelles, des consonnes sourdes, sonores et du /ʁ/ sont rapportés dans le tableau 5.3 pour le corpus CTS. Contrairement à ce que nous avions observé sur le corpus PFC, les chiffres varient très peu d'une variété à l'autre, y compris dans le cas de l'Alsace. Il faut noter le très faible taux de voisement des voyelles, pour les trois variétés : la parole téléphonique semble avoir entraîné des problèmes de détection de F_0. Les consonnes sonores sont également peu voisées, et les consonnes sourdes sont presque complètement non-voisées.

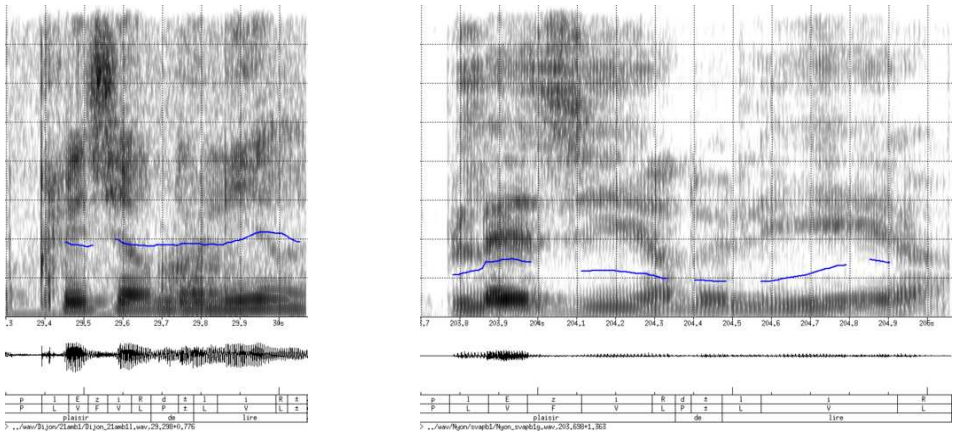
	% voisement moyen				durée moyenne (ms)			
	voyelles	C sourdes	C sonores	/ʁ/	voyelles	C sourdes	C sonores	/ʁ/
Standard	62	7	32	39	78	84	71	69
Sud	64	7	33	39	77	89	76	74
Alsace	68	6	34	37	87	90	74	72

TABLEAU 5.3. Taux de voisement moyen et durée moyenne (en ms) des voyelles, des consonnes sourdes, sonores et du /ʁ/ pour le corpus CTS.

Les mesures sur le corpus CTS ne confirment pas les tendances observées sur le corpus PFC. Comme les chiffres obtenus semblent plutôt indiquer un problème dans l'extraction des paramètres sur la parole téléphonique, nous nous sommes résolue à ne pas présenter les taux de consonnes entièrement voisées ou non-voisées pour ce corpus.

5.3 Aspects prosodiques

Dans cette section, nous utiliserons la durée, la F_0 et l'intensité, et notamment leurs variations, pour tenter de capter, à travers elles, des variations prosodiques systématiques entres nos variétés de français. La prosodie, modélisée automatiquement, a déjà été mise à profit pour identifier différentes langues ainsi que des dialectes arabes [Rouas, 2007], mais nous n'avons pas connaissance d'études similaires portant sur des variétés de français. Il semble pourtant que la prosodie varie selon les accents : la figure 5.1 montre des contours de F_0 différents pour la même suite de mots prononcée par une locutrice de Dijon et un locuteur de Nyon du corpus PFC.



(a) Dijon (b) Nyon

FIGURE 5.1. Variation de F_0 : la suite de mots « plaisir de lire » prononcée par une locutrice de Dijon et un locuteur de Nyon.

De rares travaux traitent des particularités prosodiques des accents du français [Carton *et al.*, 1991] : l'allongement des voyelles noyaux des syllabes pénultième et finale a été bien étudié pour le français de Belgique (en particulier de Liège) [Hambye et Simon, 2004]. Concernant la prosodie du français d'Alsace, [Vajta, 2002] fait l'observation suivante : *l'accent tonique germanique, situé en général sur la première syllabe, contraste fortement avec l'accent tonique français, placé sur la syllabe finale*. Pour le français de Suisse romande, [Singy, 1995 1996] note *une certaine résistance au schéma accentuel et donc au schéma intonatif du français standard, une tendance à l'accentuation de la première syllabe d'un mot bisyllabique, qui entraîne généralement une montée de la courbe mélodique*. Cette variété suisse présente pour d'autres un mouvement mélodique plus marqué en pénultième que le français standard [Grosjean *et al.*, 2007] ou une *élévation de la voix sur la syllabe précédant immédiatement la tonique* [Métral, 1977] (trait d'après le dernier auteur partagé avec d'autres variétés). Ces descriptions manquent de précision et ne nous permettent pas de définir des indices caractérisant telle ou telle variété de français — d'autant que l'accentuation initiale peut également s'observer en français standard [Léon, 1993; Welby, 2007]. La description de Singy, par exemple, peut suggérer une accentuation initiale, mais ne permet pas de la distinguer d'un patron mélodique spécifique sur la syllabe

pénultième. C'est pourquoi nous allons tester différentes hypothèses linguistiques sur nos corpus.

Nous avons calculé pour chaque phonème une valeur de F_0 et d'intensité en moyennant les valeurs définies sur l'intervalle temporel du phonème. Nous nous intéresserons à des différences de durée, de F_0 et d'intensité. L'utilisation de différences présente selon nous plusieurs avantages : nous prenons plus d'indépendance par rapport aux valeurs absolues, qui dans le cas de nos paramètres peuvent être variables. En prenant en compte deux phonèmes, nous couvrons un empan temporel plus important et nous intégrons un aspect dynamique. Pour chacun des deux corpus étudiés, nous présenterons d'abord une répartition générale des différences de durée, de F_0 et d'intensité, avant de nous intéresser plus particulièrement aux configurations prosodiques en syllabe initiale d'une part, pénultième et finale d'autre part.

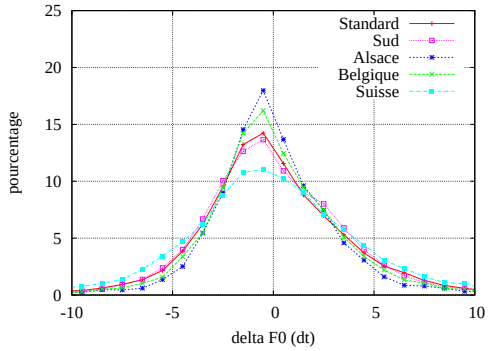
5.3.1 Description du corpus PFC

Dans la suite de ce travail, nous allons étudier la répartition des différences de chacun des trois paramètres sur des patrons spécifiques que nous décrirons dans la section suivante. Avant de passer à ces cas particuliers, nous allons examiner la répartition générale de chacune des trois différences de F_0, d'intensité et de durée. Nous avons calculé trois deltas entre des voyelles consécutives i et $i+1$: la figure 5.2 donne la répartition des $\Delta F_0 = F_{0\,i+1} - F_{0\,i}$, celle des $\Delta\text{intensité} = \text{intensité}_{i+1} - \text{intensité}_i$, et enfin celle des $\Delta\text{durée} = \text{durée}_{i+1} - \text{durée}_i$. Les courbes se lisent de la manière suivante : chaque point correspond à des valeurs cumulées. À titre d'exemple, pour la courbe (e), le point d'abcisse 0,015 indique le pourcentage de $\Delta\text{durée}$ compris entre 0 (inclus) et 0,03 ms, la valeur 0 étant comprise dans le plus petit intervalle positif de chaque courbe. Pour chaque courbe, la zone gauche (par rapport à 0) correspond à des descentes de F_0, d'intensité ou dans le cas des durées à une accélération ; la zone droite correspond à des montées ou des ralentissements.

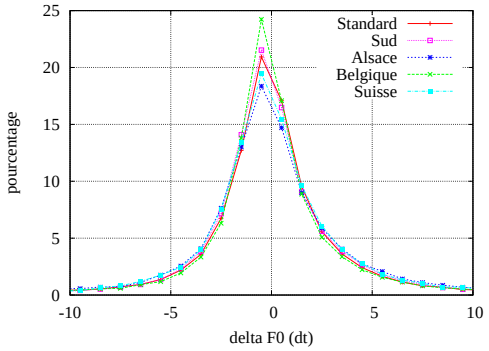
Les distributions que nous avons tracées montrent très peu de décalage entre les courbes des différentes variétés. La remarquable stabilité des résultats s'oppose aux décalages que nous avions pu observer au chapitre précédent avec les valeurs de formants et peut s'interpréter par l'utilisation de différences plutôt que de valeurs absolues. Les observations que nous pouvons faire à propos de ces courbes ne concernent pas les variétés régionales, mais plutôt des caractéristiques générales de la parole de notre corpus.

Pour les ΔF_0, la différence entre styles de parole est plus marquée que les différences entre variétés. F_0 est plus variable en parole spontanée, où les proportions autour de 0 cumulent entre 35 et 40 %, qu'en lecture, où les proportions restent globalement entre 20 et 30 %. Le maximum des courbes de ΔF_0 est à gauche par rapport au 0 : de légères descentes de F_0 sont donc le cas le plus fréquent. C'est également la cas pour les $\Delta\text{intensité}$, pour lesquels aucune différence notable n'apparaît selon le style de parole. Le maximum des courbes de $\Delta\text{durée}$ est à droite par rapport à 0 : il y a plus de petits ralentissements que de petites accélérations. Ces observations sont valables pour les deux styles de parole.

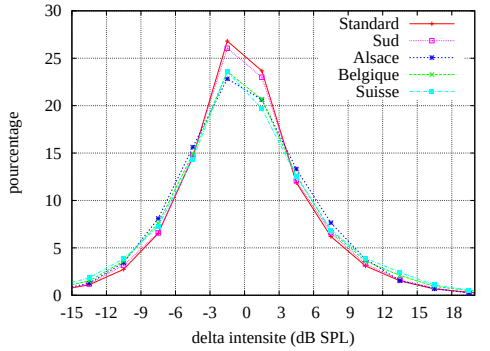
Nous présenterons également par la suite les durées de certains phonèmes. Le tableau 2.4,



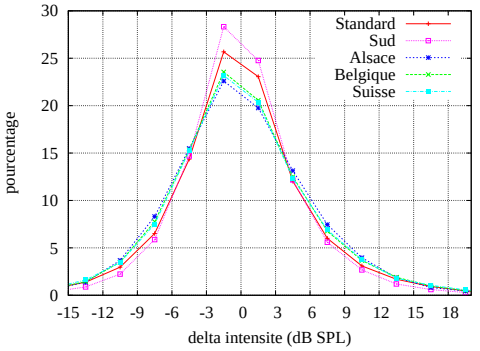
(a) ΔF_0, lecture



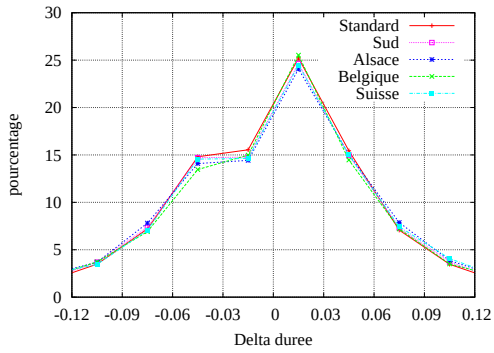
(b) ΔF_0, parole spontanée



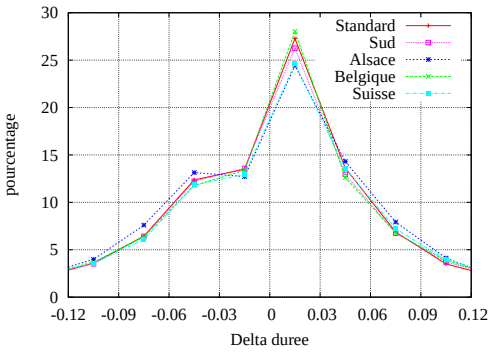
(c) Δintensité, lecture



(d) Δintensité, parole spontanée



(e) Δdurée, lecture



(f) Δdurée, parole spontanée

FIGURE 5.2. Répartition des ΔF_0 (en demi-tons), Δintensité (en dB SPL) et Δdurée (en s) pour le corpus PFC, en lecture et en parole spontanée.

p. 33 donne la durée moyenne de tous les phonèmes et le tableau 5.1, p. 87 celle des voyelles pour les différentes variétés de français. Les Suisses présentent un débit plutôt lent par rapport aux autres variétés de français, notamment en parole spontanée. Les Alsaciens ont des voyelles plus longues en lecture.

5.3.2 Accentuation initiale dans le corpus PFC

La présentation des mesures globales, nous l'avons vu, ne nous permet d'observer que très peu de différences entre les variétés. Nous pensons que l'accent peut se traduire, non pas nécessairement comme un trait ou une marque permanente, mais par des réalisations ciblées, inattendues par rapport au standard, à certains endroits plus ou moins aléatoires dont l'accentuation initiale pourrait faire partie en donnant lieu à une accentuation stylistique, mais également à une accentuation marquant l'origine régionale.

Pour caractériser l'accentuation initiale, nous avons examiné les suites clitique non-clitique (c'est-à-dire un mot-outil suivi d'un mot porteur de sens), qui sont de bons candidats pour recevoir un accent sur le non-clitique [Adda-Decker, 2006]. Nous avons retenu comme clitiques une trentaine de monosyllabes parmi les mots les plus fréquents du français (*le, la, les, un, une, de, du, des, en, pour*…), lesquels assurent une bonne couverture du contexte étudié [Boula de Mareüil *et al.*, 2008]. Pour les non-clitiques, nous nous sommes restreinte aux polysyllabes (sans compter le schwa final qui peut être prononcé).

Nous avons mesuré les variations de F_0, d'intensité et de durée entre la voyelle noyau du clitique et celle de la première syllabe du mot suivant, ainsi que la durée de l'attaque du non-clitique. Tous ces calculs sont illustrés par la figure 5.3, où l'on considère l'exemple de la suite de mots *pour trouver* : les noyaux vocaliques sont les deux /u/, et l'attaque est constituée de la séquence /tʁ/.

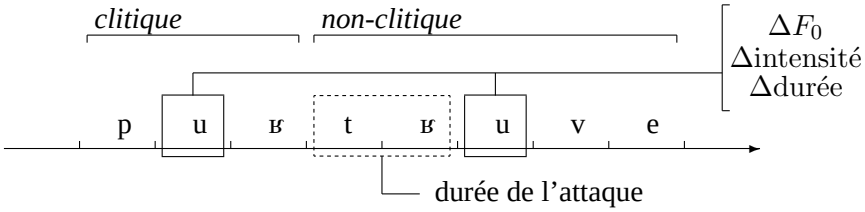


FIGURE 5.3. Calculs liés à l'accentuation initiale.

Différences de F_0

Nous avons calculé le pourcentage d'occurrences compris entre différents seuils pour les différences de fréquence fondamentale $\Delta F_0 = F_{0\,\mathrm{non-clitique}} - F_{0\,\mathrm{clitique}}$ en demi-tons (dt). Les résultats cumulés sont présentés dans le tableau 5.4 pour des pourcentages de ΔF_0 supérieurs à des seuils de 1, 2 et 3 demi-tons (3 demi-tons correspondant au seuil à partir duquel une proéminence prosodique peut jouer un rôle linguistique selon [' t Hart *et al.*,

1991]) — un ΔF_0 positif plus ou moins grand correspond à une montée de F_0 plus ou moins importante. Les nombres d'occurrences de contextes à partir desquels les pourcentages sont exprimés sont également donnés, pour la lecture et la parole spontanée. Les résultats sont donnés pour les polysyllabes et les mots de trois syllabes et plus. En effet, les polysyllabes sont en majorité composés de bisyllabes, pour lesquels les syllabes initiales et pénultièmes sont confondues. Les mots de trois syllabes et plus permettent de ne pas inclure les syllabes pénultièmes dans le calcul.

		polysyllabes				trisyllabes +			
		#occ	% Δ >1 dt	% Δ >2 dt	% Δ >3 dt	#occ	% Δ >1 dt	% Δ >2 dt	% Δ >3 dt
lecture	Standard	2035	55	41	29	533	68	54	41
	Sud	1950	53	40	25	510	68	53	37
	Alsace	414	35	21	14	107	38	24	16
	Belgique	1278	46	30	20	333	59	44	28
	Suisse	435	**76**	**60**	**46**	117	**81**	**68**	**58**
spontané	Standard	7242	26	16	11	1309	26	16	10
	Sud	6078	22	13	8	1150	21	12	7
	Alsace	999	24	14	10	144	26	15	11
	Belgique	2498	18	10	7	439	17	9	6
	Suisse	1647	**40**	**25**	**16**	306	**48**	**30**	**19**

TABLEAU 5.4. Nombre d'occurrences et pourcentage de ΔF_0 (en demi-tons) clitique non-clitique (polysyllabique et au moins trisyllabique) supérieur à un seuil donné.

Les résultats vont dans le même sens en considérant tous les polysyllabes ou en restreignant l'analyse aux non-clitiques d'au moins trois syllabes pour lesquels le nombre de cas examinés est bien entendu moins élevé. Seuls les Suisses présentent des taux plus élevés dans la seconde configuration que dans la première.

Quels que soient le style et le seuil considérés, les pourcentages plus élevés pour les Suisses que pour les autres locuteurs, ainsi que le montre le tableau 5.4, traduisent une proportion plus importante de montées de F_0 pour les Suisses. Le phénomène s'observe clairement dans la courbe de distribution (voir la figure 5.4 pour la parole spontanée). Sur les polysyllabes, il est toutefois plus marqué dans le texte lu : l'écart avec le français standard atteint 21 %, contre 14 % pour la parole spontanée. Les valeurs obtenues pour les locuteurs du Sud, d'Alsace et de Belgique sont au contraire plus basses que celles du français standard. Les locuteurs alsaciens présentent un peu moins de ΔF_0 supérieurs à chacun des seuils en lecture, mais cette tendance ne se retrouve pas en parole spontanée. La courbe de distribution montre également que les locuteurs alsaciens semblent présenter le phénomène inverse des Suisses.

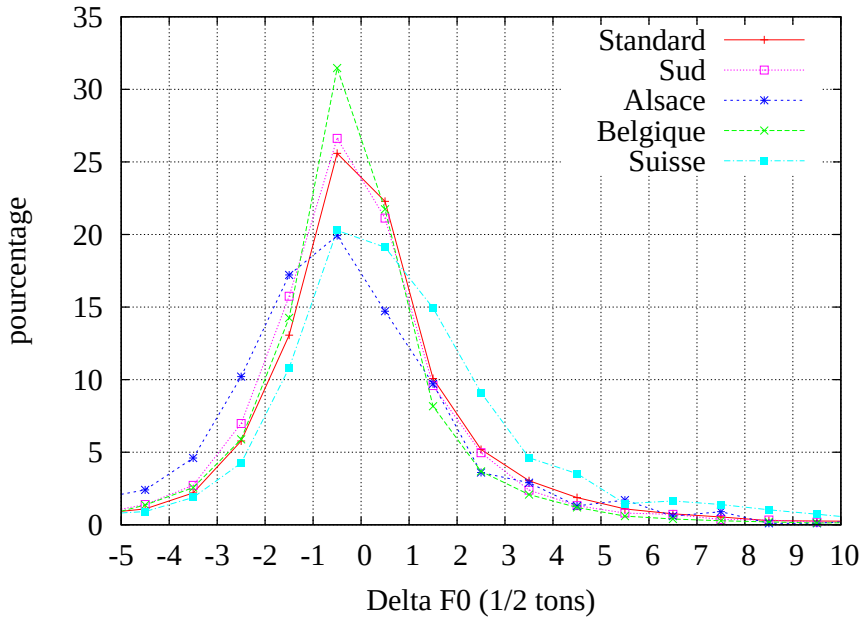


FIGURE 5.4. Distribution des ΔF_0 (en demi-tons) clitique non-clitique polysyllabique pour la parole spontanée.

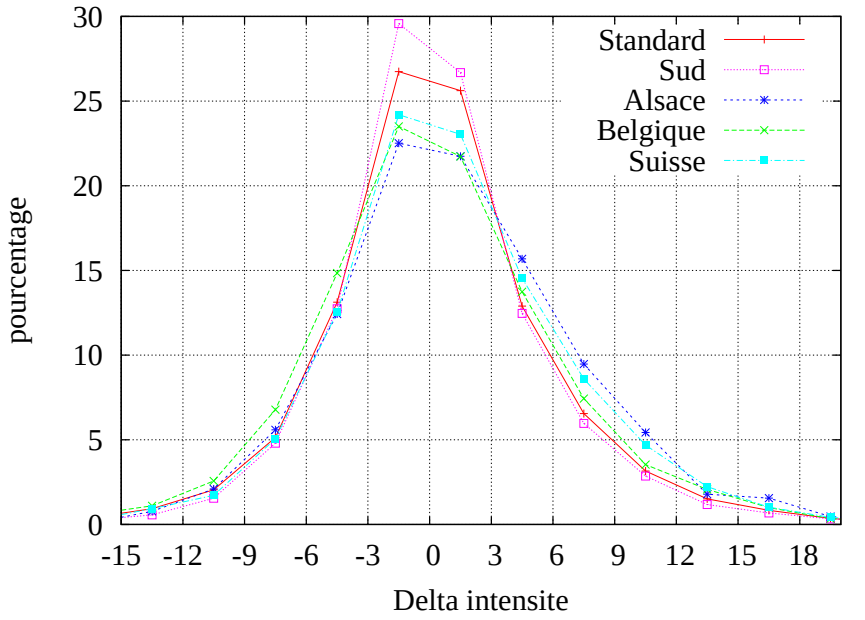


FIGURE 5.5. Distribution des Δintensité (en dB SPL) clitique non-clitique polysyllabique pour la parole spontanée.

Différences d'intensité

De la même manière que pour F_0, nous avons calculé les différences d'intensité $\Delta\text{intensité} = \text{intensité}_{\text{non-clitique}} - \text{intensité}_{\text{clitique}}$ en dB SPL. Comme pour F_0, un $\Delta\text{intensité}$ positif correspond à une montée de l'intensité, un $\Delta\text{intensité}$ négatif à une diminution. Un accent initial peut s'accompagner d'une augmentation de l'intensité : nous nous attendons dans ce cas à trouver plus de $\Delta\text{intensité}$ positifs.

Les pourcentages cumulés de $\Delta\text{intensité}$ supérieurs à des seuils de 0 et 3 dB SPL sont présentés dans le tableau 5.5. En lecture, ce sont les Suisses, suivis des Alsaciens, qui présentent les taux les plus élevés de $\Delta\text{intensité}$ supérieurs à chaque seuil ; pour la parole spontanée, ce sont les Alsaciens suivis des Suisses avec des écarts faibles. Les locuteurs du Sud et les Belges montrent des pourcentages assez proches de ceux du français standard, légèrement plus ou moins élevés selon la configuration. La figure 5.5 montre la distribution des $\Delta\text{intensité}$ pour la parole spontanée : les Alsaciens et les Suisses présentent une courbe légèrement décalée vers la droite.

		polysyllabes			% trisyllabes +		
		#occ	% Δ >0 dB	% Δ >3 dB	#occ	% Δ >0 dB	% Δ >3 dB
lecture	Standard	2085	57	27	544	45	15
	Sud	1587	59	30	414	49	18
	Alsace	450	**61**	**34**	113	**56**	**24**
	Belgique	1356	56	30	350	43	14
	Suisse	443	**76**	**47**	118	**69**	**32**
spontané	Standard	8740	51	26	1576	51	25
	Sud	7252	50	24	1351	51	23
	Alsace	1288	**56**	**35**	174	**58**	**34**
	Belgique	3820	50	29	667	51	26
	Suisse	2153	**55**	**32**	400	**55**	**32**

TABLEAU 5.5. Nombre d'occurrences et pourcentage de $\Delta\text{intensité}$ clitique non-clitique (polysyllabique et au moins trisyllabique) supérieur à un seuil donné.

Il est intéressant de remarquer que les deux variétés pour lesquelles nous avons observé plus de $\Delta\text{intensité}$ positifs, à savoir la Suisse et l'Alsace, sont également celles qui présentaient plus de variation de F_0 : nettement plus de montées pour la Suisse, et plus de descentes pour l'Alsace.

Différences de durée

La différence $\Delta\text{durée} = \text{durée}_{\text{non-clitique}} - \text{durée}_{\text{clitique}}$ a de même été calculée entre voyelles. Ici, un $\Delta\text{durée}$ positif correspond à un allongement de la voyelle de la syllabe initiale du mot non-clitique par rapport à celle du clitique, soit un ralentissement ; un $\Delta\text{durée}$ négatif correspond à une accélération.

		polysyllabes			% trisyllabes +		
		#occ	% Δ >0 ms	% Δ >20 ms	#occ	% Δ >0 ms	% Δ >20 ms
lecture	Standard	2085	56	32	544	57	30
	Sud	1587	61	35	414	62	37
	Alsace	450	**62**	**40**	113	**67**	**43**
	Belgique	1356	62	36	350	62	35
	Suisse	443	55	29	118	51	23
spontané	Standard	8740	55	28	1576	52	26
	Sud	7252	55	29	1351	53	27
	Alsace	1288	**64**	**41**	174	**57**	**39**
	Belgique	3820	57	30	667	51	26
	Suisse	2153	57	33	400	53	31

TABLEAU 5.6. Nombre d'occurrences et pourcentage de Δdurée clitique non-clitique (polysyllabique et au moins trisyllabique) supérieur à un seuil donné.

Les résultats pour les polysyllabes et les mots d'au moins trois syllabes sont présentés dans le tableau 5.6. Environ un tiers des patrons examinés montrent un Δdurée >20 ms et plus de la moitié des patrons un Δdurée > 0 ms dans toutes les conditions : il y a plus d'occurrence pour lesquelles la voyelle du clitique est plus courte que celle de l'initiale du non-clitique subséquent.

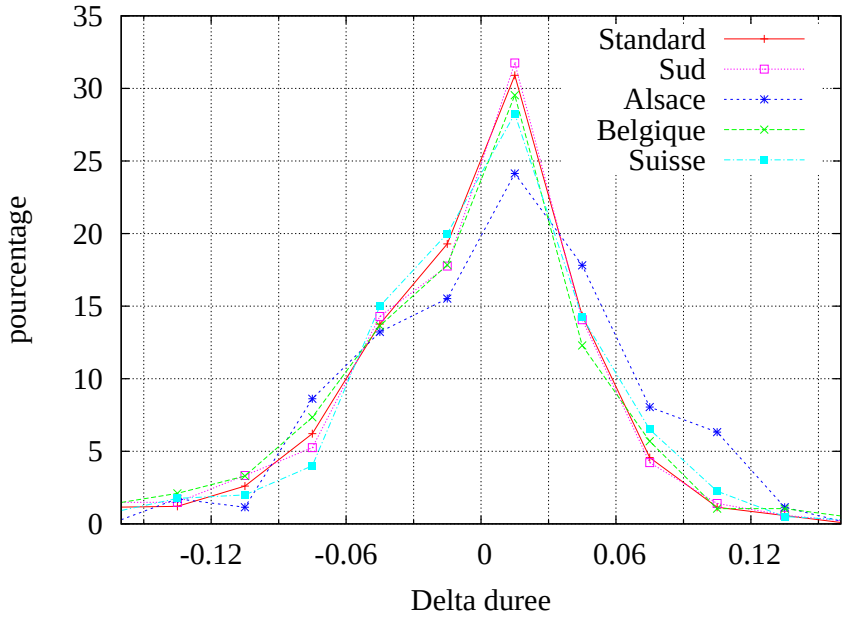


FIGURE 5.6. Distribution des Δdurée (en s) clitique non-clitique au moins trisyllabique pour la parole spontanée.

Les Alsaciens présentent le plus fort pourcentage de $\Delta\text{durée}$ supérieur à chacun des deux seuils proposés (0 et 20 ms) en parole spontanée et, dans une moindre mesure, en lecture. La même tendance s'observe avec un seuil de 30 ms. Si les non-clitiques sont restreints aux mots d'au moins trois syllabes, ce qui nous permet de faire la part entre syllabes initiales et pénultièmes, la différence entre les locuteurs alsaciens et les autres est encore plus marquée ; c'est ce qui apparaît sur la figure 5.6 correspondant à la parole spontanée.

Les durées moyennes des noyaux vocaliques du clitique et de la syllabe initiale du non-clitique subséquent sont présentées dans la tableau 5.7 (voir section suivante). La durée moyenne du noyau du clitique est indiquée bien qu'elle n'apporte pas beaucoup d'information et permet simplement de constater qu'il y a entre régions peu de variation, surtout en parole spontanée. La première voyelle du mot plein qui suit se comporte quant à elle comme prévu d'après les calculs de $\Delta\text{durée}$: les Alsaciens ont les durées moyennes les plus longues, ce qui n'est pas seulement imputable à un débit plus lent car les Suisses montrent un comportement très différent : en lecture, les durées moyennes du mot plein sont comparables à celles des autres locuteurs, elles s'allongent en spontanée mais sans atteindre celles des Alsaciens.

Allongement de l'attaque

D'après [Jankowski *et al.*, 1999; Mertens, 1993], l'allongement de l'attaque est un corrélat de l'accentuation initiale. Nous avons mesuré la durée moyenne des attaques des mots pleins suivant un clitique : les résultats sont donnés dans le tableau 5.7 (le cas échéant restreints aux attaques simples, *i.e.* composées d'une seule consonne). Le nombre de contextes considérés dans chaque cas se déduit du chiffre indiqué dans le tableau 5.6, auquel il faut soustraire les suites clitique non-clitique ne comportant pas d'attaques (soit environ 15 % des cas). Les contextes avec attaque simple, eux, représentent environ deux tiers des nombres d'occurrences donnés tableau 5.6.

En lecture, les attaques simples sont en moyenne plus longues que les voyelles qui les entourent. Les attaques sont plus longues pour les mots d'au moins trois syllabes, tandis que les durées des voyelles restent similaires. En parole spontanée, les attaques simples ont une durée moyenne équivalente à celles de la voyelle du clitique. Les durées de ces deux phonèmes sont supérieures à celle de la voyelle du non-clitique. Les deux styles de parole présentent des comportements différents.

Les Alsaciens, qui réalisaient les voyelles les plus longues en syllabe initiale de polysyllabe suivant un clitique, ne semblent pas allonger l'attaque. Ce sont les Suisses qui montrent des attaques plus longues.

Ainsi, l'accentuation initiale chez nos locuteurs suisses semble se manifester par une augmentation de F_0 et un léger allongement de l'attaque, indices bien décrits dans la littérature [Jankowski *et al.*, 1999; Mertens, 1993]. Ils présentent également une légère augmentation de l'intensité, en commun avec nos locuteurs alsaciens, qui montrent par ailleurs plutôt un allongement de la voyelle initiale de polysyllabes. Que ce trait soit interprété en terme d'accentuation initiale ou non, nos résultats suggèrent que des indices acoustiques différencient la prosodie du français parlé en Alsace et en Suisse romande. La Belgique ne montre guère

		polysyllabes			% trisyllabes +		
		Vc	attaque [simple]	Vnc	Vc	attaque [simple]	Vnc
lecture	Standard	72	88 [80]	72	70	105 [92]	70
	Sud	69	94 [84]	73	65	109 [95]	68
	Alsace	76	92 [82]	**83**	73	106 [91]	**81**
	Belgique	68	89 [81]	72	64	102 [86]	65
	Suisse	69	**97 [90]**	70	63	**114 [100]**	60
spontané	Standard	76	91 [78]	63	78	89 [77]	61
	Sud	78	88 [76]	64	77	86 [75]	61
	Alsace	78	86 [75]	**79**	84	89 [73]	**82**
	Belgique	76	90 [75]	65	76	87 [74]	60
	Suisse	78	**98 [85]**	71	77	**97 [85]**	68

TABLEAU 5.7. Durée moyenne en ms du noyau du clitique (Vc), du noyau (Vnc) et de l'attaque de la première syllabe du non-clitique, polysyllabe ou au moins trisyllabe. Entre crochets, seules les attaques simples sont prises en compte.

de spécificité par rapport au français standard, au regard de ces paramètres, mais d'autres faits prosodiques peuvent amener une différenciation. C'est ce que nous étudierons dans la section suivante.

5.3.3 Syllabes précédent une pause dans le corpus PFC

Nous avons examiné, en dernier lieu, le comportement des syllabes pénultièmes et finales avant une pause (nous avons considéré comme pauses les silences de plus de 50 ms, ce qui donne des segments de parole inter-pauses de 1,6 s en moyenne). Le pourcentage d'avant-dernières voyelles plus longues que les finales (schwa exclu) permet de saisir une forme d'allongement pénultième. Nous nous attendons dans le cas standard à observer un ralentissement avant les pauses, avec des syllabes de plus en plus longues, ce qui implique que la durée de la syllabe finale est supérieure à celle de la pénultième.

La figure 5.7 illustre les voyelles considérées dans les calculs de cette section pour le mot *tombola* : nous nous intéressons aux différences de durée entre l'antépénultième /õ/, la pénultième /o/ et la finale /a/.

Après avoir également observé les variations tant de F_0 que d'intensité entre les syllabes pénultième et finale, nous n'avons pas pu dégager de tendance marquée selon la région. En conséquence, nous présenterons uniquement les résultats concernant les variations de durée.

Le pourcentage de $\Delta\text{durée}_\text{f} = \text{durée}_\text{pénultième} - \text{durée}_\text{finale}$ positif est donné pour chaque région dans le tableau 5.8, avec les durées moyennes des voyelles noyaux des deux syllabes précédant une pause. Un $\Delta\text{durée}_\text{f}$ positif traduit ici un allongement de la voyelle noyau de la pénultième syllabe par rapport à celle de la syllabe finale. Le nombre d'occurences

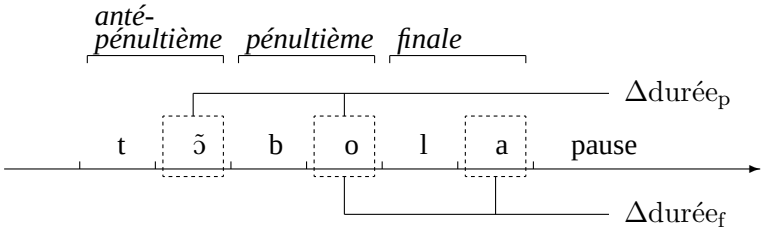


FIGURE 5.7. Calculs liés aux syllabes précédant une pause.

examinées, relativement peu élevé pour les mots d'au moins trois syllabes, est également indiqué.

En lecture, les pourcentages sont assez proches entre les régions, que ce soit pour les mots d'au moins trois syllabes ou les polysyllabes. Considérer les mots d'au moins trois syllabes évite de confondre allongements initial et pénultième, qui pourraient être caractéristiques de variétés différentes. Les chiffres diffèrent davantage sur la parole spontanée : les locuteurs belges (de Liège, Gembloux et Tournai) réalisent plus de $\Delta\text{durée}_\text{f} = \text{durée}_\text{pénultième} - \text{durée}_\text{finale} > 0$ ms que ceux de français standard (différence de 10 % au moins), ce qui correspond à des pénultièmes plus longues par rapport aux finales, tandis que les Alsaciens et les Suisses en réalisent moins. Des résultats similaires sont obtenus en normalisant la durée de chaque voyelle par sa durée moyenne calculée sur l'ensemble du corpus.

		polysyllabes				% trisyllabes +			
		#occ	% Δ >0 ms	dur. pén.	dur. fin.	#occ	% Δ >0 ms	dur. pén.	dur. fin.
lecture	Standard	1162	23	83	148	433	25	82	144
	Sud	1111	22	87	158	400	19	79	168
	Alsace	289	27	86	134	110	21	77	134
	Belgique	932	**29**	86	142	334	**27**	82	146
	Suisse	326	23	90	**154**	114	25	93	**153**
spontané	Standard	3302	32	68	123	790	26	66	129
	Sud	2986	30	70	134	874	25	65	139
	Alsace	472	39	83	113	111	26	64	112
	Belgique	2176	**40**	72	122	494	**37**	73	133
	Suisse	1077	30	82	**154**	252	23	79	**166**

TABLEAU 5.8. Nombre d'occurrences et pourcentage de $\Delta\text{durée}_\text{f}$ pénultième-finale positif avant une pause, ainsi que les durées moyennes des deux dernières voyelles en ms.

Les voyelles nasales sont connues pour être plus longues que les orales (112 *vs* 80 ms en moyenne, dans notre corpus). De fait, l'écart se creuse légèrement entre locuteurs belges et français standard (atteignant 15 %) en considèrant les polysyllabes dont la voyelle pénultième est une nasale.

En durées moyennes, les Suisses ont des voyelles finales assez longues. En comparaison avec le français standard, leurs voyelles pénultièmes sont aussi plus longues (avec une plus petite marge), ce qui explique les valeurs plus faibles de $\Delta\text{durée}_\text{f}$. Ce dernier résultat n'étaye pas une tendance à l'allongement pénultième en Suisse romande, mais amène à poser l'hypothèses que les Suisses allongeraient à la fois l'avant-dernière et la dernière voyelle.

Pour vérifier cette hypothèse, nous avons calculé le pourcentage de voyelles pénultièmes plus longues que la voyelle antépénultième la précédant. Pour ce calcul, nous n'avons pris en compte que les mots d'au moins trois syllabes, les disyllabes ne comportant pas de syllabe antépénultième. Les résultats présentés dans le tableau 5.9 montrent que les Suisses réalisent plus de $\Delta\text{durée}_\text{p} = \text{durée}_\text{pénultième} - \text{durée}_\text{antépénultième} > 0$ ms. La durée moyenne de leur antépénultième est parmi les plus courtes avec celles des Belges, et ils présentent moins d'antépénultièmes d'une durée supérieure à 70 ms que les autres variétés.

Nous constatons sans surprise que les Alsaciens ont des antépénultièmes assez longues : dans la majorité des cas, cette syllabe est également l'initiale (la plupart des mots étudiés ici sont trisyllabiques). Ce résultat est cohérent avec l'allongement de la syllabe initiale décrit précédemment. Les Belges ont, en parole spontanée, le plus fort pourcentage de pénultièmes plus longues que les antépénultièmes, ce qui conforte l'hypothèse de l'allongement pénultième pour ces locuteurs.

		% trisyllabes +			
		#occ	% Δ >0 ms	dur. ant.	% dur. ant. >70 ms.
lecture	Standard	433	66	70	37
	Sud	400	60	74	37
	Alsace	110	55	**78**	39
	Belgique	334	**71**	66	28
	Suisse	114	**81**	65	26
spontané	Standard	790	65	61	27
	Sud	874	59	64	29
	Alsace	111	55	**72**	41
	Belgique	494	**70**	62	27
	Suisse	252	**69**	65	26

TABLEAU 5.9. Pourcentage de $\Delta\text{durée}_p$ (pénultième - antépénultième) positifs avant une pause, durée moyenne des voyelles antépénultièmes (en ms) et pourcentage de voyelles antépénultièmes de durée > à 70 ms dans des mots d'au moins 3 syllabes.

En examinant le comportement des trois dernières syllabes avant une pause, nos résultats suggèrent quelques comportements spécifiques pour certaines variétés de français : les Belges allongent la syllabe pénultième par rapport à l'antépénultième et à la finale, tandis que les Suisses allongent les deux dernières syllabes par rapport à l'antépénultième. Il faut toutefois noter que ces tendances en matière de patron final sont moins nettement marquées que celles que nous avons mesurées en matière de patron initial.

5.3.4 Description du corpus CTS

De la même manière que pour le corpus PFC, avant de nous intéresser à des patrons particuliers, nous allons étudier la répartition des deltas sur toutes les syllabes consécutives pour l'ensemble du corpus CTS. Nous n'observerons ici cette répartition que pour les $\Delta\text{durée}$, le seul paramètre présenté par la suite étant $\Delta\text{durée}$ en initiale. En effet, un certain nombre de paramètres que nous avions présentés pour le corpus PFC n'ont pas révélé de différences marquées entre les variétés de français standard, du Sud et d'Alsace une fois appliqués au corpus CTS. Cette absence de différence n'est pas surprenante puisque ces calculs, appliqués sur le corpus PFC, avaient plutôt mis en évidence des différences pour la Belgique et la Suisse, qui ne sont pas représentées dans le corpus CTS (limité aux frontières de la France). C'est pourquoi nous ne présenterons dans cette section ni les ΔF_0, ni les $\Delta\text{intensité}$, ni l'allongement de l'attaque pour l'accentuation initiale. Nous avions remarqué une légère augmentation de l'intensité sur la syllabe initiale en Alsace qui n'a pas été retrouvée ici. Nous avons de plus rencontré lors du traitement des syllabes précédant une pause des difficultés en partie dues à un effet du tour de parole lié à l'interaction téléphonique (limitée uniquement à la voix). Les résultats obtenus n'ont pas non plus fait apparaître de différence entre les variétés étudiées et par conséquent nous ne les avons pas présenté ici. Du reste, comme ces paramètres, de la même manière que les précédents, caractérisent plutôt les Belges et les Suisses, l'absence de différence n'est pas surprenante.

La distribution des $\Delta\text{durée} = \text{durée}_{i+1} - \text{durée}_i$ est donnée par la figure 5.8 pour les trois variétés de français du corpus CTS : standard, Sud et Alsace. Comme pour le corpus PFC, nous n'observons pas de décalage entre les courbes présentées pour les différentes variétés. Le pic se trouve à droite du 0 et signifie que les petits ralentissements sont plus fréquents que les petites accélérations. Le fait que la profil de cette courbe soit très proche de celui que nous avions observé pour le corpus PFC (figure 5.2, page 92) confirme l'hypothèse selon laquelle les caractéristiques de cette distribution sont indépendantes de la variété de français concernée.

Rappelons que les durées moyennes des phonèmes sont données dans le tableau 2.5, p. 34, et celles des voyelles dans le tableau 5.3, p. 89.

5.3.5 Accentuation initiale dans le corpus CTS

Pour les raisons données dans la description du corpus, nous nous cantonnerons à présenter ici les $\Delta\text{durée} = \text{durée}_{\text{non-clitique}} - \text{durée}_{\text{clitique}}$ calculés entre voyelles (pour une illustration du calcul, voir la figure 5.3, p. 93). Rappelons que les $\Delta\text{durée}$ positifs correspondent à des ralentissements, la voyelle noyau du non-clitique étant allongée par rapport à celle du clitique.

Le nombre de $\Delta\text{durée}$ supérieurs à des seuils donnés est présenté dans le tableau 5.10, pour les polysyllabes et les mots d'au moins trois syllabes. Les Alsaciens montrent le plus fort pourcentage de $\Delta\text{durée}$ supérieurs à chacun des deux seuils proposés. Dans le corpus PFC, la tendance était plus marquée sur les mots d'au moins trois syllabes (voir tableau 5.6) : c'est le cas également ici, quoique plus légèrement. La tendance des Alsaciens à allonger la voyelle de la syllabe initiale est illustrée par la figure 5.9 pour les mots d'au moins trois

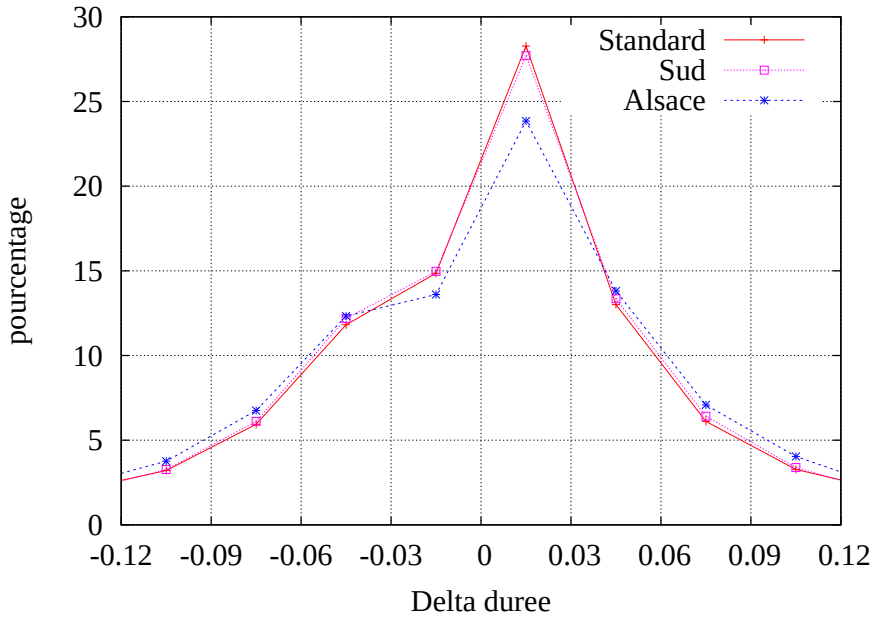


FIGURE 5.8. Répartition des Δdurée (en s) pour le corpus CTS.

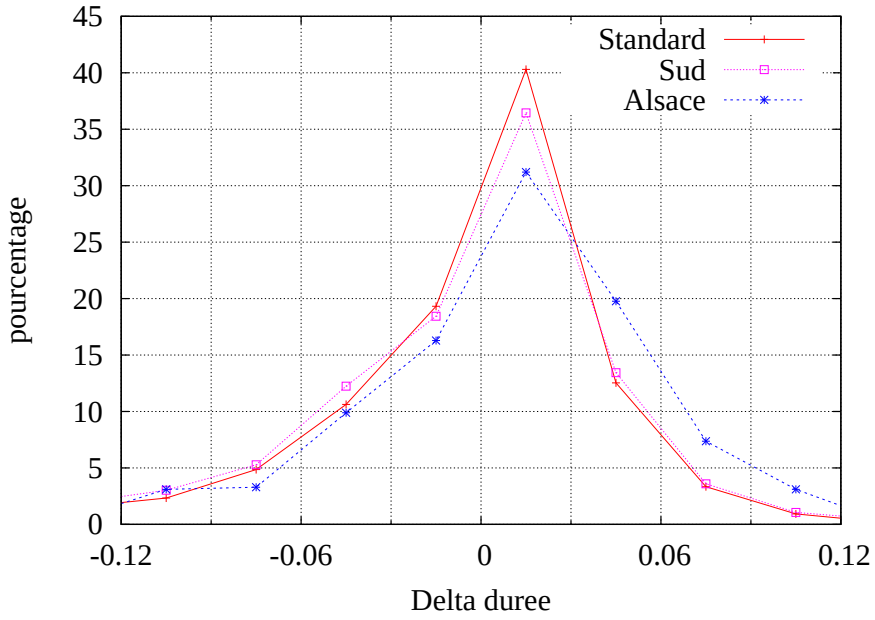


FIGURE 5.9. Distribution des Δdurée (en s) clitique non-clitique au moins trisyllabique pour le corpus CTS.

syllabes. La courbe de distribution des $\Delta\text{durée}$ est effectivement décalée vers la droite pour ces locuteurs, de manière très similaire à ce que nous avons pu observer sur le corpus PFC (figure 5.6, page 97).

	polysyllabes			% trisyllabes +		
	#occ	% Δ >0 ms	% Δ >20 ms	#occ	% Δ >0 ms	% Δ >20 ms
Standard	19 755	58	27	3 566	57	17
Sud	12 792	58	28	2 528	56	18
Alsace	2 726	63	38	516	63	31

TABLEAU 5.10. Nombre d'occurrences et pourcentage de $\Delta\text{durée}$ clitique non-clitique (polysyllabique et au moins trisyllabique) supérieur à un seuil donné.

5.4 Conclusion

Nous avons dans ce chapitre étudié les paramètres de durée, de F_0 et d'intensité, d'une part pour certaines consonnes, telles que les plosives, les fricatives et le /ʁ/ et d'autre part pour certains patrons prosodiques.

Au niveau des consonnes, nous n'avons pas trouvé de variation de durée liée à l'origine géographique des locuteurs. En ce qui concerne les taux de voisement, les résultats sont mitigés. Les variétés d'Alsace, de Belgique et de Suisse se démarquent du français standard par un voisement moindre, tandis que les locuteurs du Sud ont un comportement très similaire à celui du français standard. Les Alsaciens présentent les taux de voisement de consonnes les plus bas : ce comportement semble concorder avec les descriptions de l'accent alsacien dans la littérature. Les Belges présentent, au moins en parole spontanée, des /ʁ/ moins voisés que ceux des autres variétés étudiées. Toutefois, des variations apparaissent également pour les voyelles, qui présentent des taux de voisement plus ou moins proche de 100 % qui peuvent être interprétées comme un indice de disparités dues aux conditions d'enregistrement dans la détection de F_0. Nous n'avons pu ni confirmer ni infirmer nos résultats sur le corpus CTS, pour lequel d'évidents problèmes de détection de F_0 se sont posés du fait du type de parole (conversations téléphoniques).

Dans le chapitre suivant, nous utiliserons une autre méthode pour analyser le comportement des consonnes et des voyelles. Nous pourrons essayer de faire la part des choses entre différences régionales et problèmes techniques et voir si les résultats vont dans le même sens.

L'extraction de la fréquence fondamentale et de l'intensité ont permis, sinon de révéler, du moins de quantifier des différences prosodiques entre variétés de français. La Suisse romande se caractérise par une tendance à l'accentuation initiale (montée de la mélodie, légère augmentation de l'intensité et allongement de l'attaque). En Alsace, l'allongement de la voyelle initiale — trait présent dans nos deux corpus — peut s'interpréter comme une accentuation initiale sous l'influence du contact des langues. Nous avons également observé

une légère augmentation de l'intensité pour les locuteurs alsaciens, mais uniquement sur le corpus PFC. L'allongement de la voyelle pénultième précédant une pause semble plutôt typique de la Belgique, tandis que les Suisses ont plutôt tendance à allonger les deux dernières syllabes avant la pause. Tous ces corrélats acoustiques apparaissent soit robustes au style (lecture ou parole spontanée), soit plus marqués sur la parole spontanée, ce qui peut s'expliquer par le fait que la parole lue est plus normée.

Les mesures que nous avons faites pour mettre en évidence des différences prosodiques sont fondées sur des patrons relativement généraux. L'examen de cas plus particuliers pourrait faire apparaître d'autres indices.

Enfin, nous aimerions utiliser les caractéristiques trouvées afin de classifier automatiquement nos variétés de français. Les indices prosodiques sont plutôt fins : trouveront-ils leur utilité dans une tâche de classification ? Le système automatique préfèrera peut-être des indices pour lesquels la différence est plus marquée. Nous étudierons cette question dans le chapitre 7.

Chapitre 6

Une approche à base de variantes de prononciation

Jusqu'ici, nous avons utilisé les frontières temporelles fournies par un alignement en phonèmes « standard » pour extraire des paramètres acoustiques du signal. Dans ce chapitre, nous allons étudier les différences phonétiques entre régions d'une nouvelle manière : nous introduirons des variantes liées aux accents régionaux dans le dictionnaire de prononciation et nous observerons les variantes choisies par le système. Ce travail repose sur la méthodologie introduite par [Adda-Decker et Lamel, 1999] afin d'étudier les variantes de prononciation en fonction de la configuration du système d'alignement. Les taux de réalisations que nous obtiendrons pour les variantes seront interprétés de manière relative entre les différentes variétés étant donné que les valeurs absolues de ces taux peuvent varier en fonction de la configuration du système.

Certains phénomènes tels que la réalisation du schwa (ou *e* muet) et des voyelles nasales peuvent difficilement être étudiés sans introduire des variantes de prononciation. L'approche adoptée dans ce chapitre nous permettra de les examiner. A contrario, l'étude de certains phénomènes peut être envisagée sous plusieurs angles : nous analyserons à l'aide des variantes de prononciation la réalisation du /ɔ/ ouvert, déjà étudiée à travers les mesures de formants, ainsi que le voisement des consonnes, déjà étudié grâce aux mesures de fréquence fondamentale. Nous pourrons ainsi déterminer si les deux approches vont dans le même sens, ce qui permettrait de corroborer les résultats établis par les analyses acoustiques.

6.1 Réalisation des voyelles nasales

La prononciation des voyelles nasales a fait l'objet d'études aussi bien pour le français standard [Martinet, 1945; Malécot et Lindsay, 1976; Malderez, 1991; Léon, 1993; Delvaux *et al.*, 2002; Montagu, 2004] que pour le français du Sud [Martinet, 1945; Walter, 1982; Thomas, 1991; Binisti et Gasquet-Cyrus, 2003; Clairet, 2005]. Les locuteurs méridionaux sont connus pour prononcer fréquemment des voyelles partiellement, voir to-

talement dénasalisées et suivies d'une consonne nasale. Cet appendice nasal s'articule au même lieu que la consonne suivante.

L'étude des voyelles nasales est complexe du point de vue acoustique. Ces voyelles se prêtent mal à l'analyse formantique : le couplage du conduit nasal et du conduit oral fait apparaître des paires de formants et anti-formants qui rendent difficile la détection du formant oral. De plus, les effets de la nasalité sont variables selon la voyelle et le locuteur. L'alignement automatique, qui permet d'aller au-delà du simple phonème et d'apparier par exemple /ɑ̃/ et [a]+[n], est donc particulièrement bien adapté à l'examen des voyelles nasales, exclues de l'analyse menée au chapitre 4.

6.1.1 Variantes proposées

Nous voulons proposer pour les voyelles nasales des variantes éventuellement dénasalisées, avec un appendice consonantique nasal optionnel. Il convient de préciser que, dans notre système d'alignement pour le français, nous ne disposons pas de modèles acoustiques pour le /ŋ/ et le /œ̃/, or la question de leur utilisation peut se poser ici.

Nous avons mené une première expérience en introduisant des *xénophones* (phonèmes issus du système phonémique d'une autre langue) pour /ŋ/. Mais que ce soit en utilisant un /ŋ/ provenant de modèles acoustiques anglais ou allemands, nous n'avons pas observé de différence notable entre les variétés standard et Sud, ce qui nous a amenée à ne pas conserver le /ŋ/ dans la suite de ce travail. La neutralisation de l'opposition entre /œ̃/ et /ɛ̃/ dans des paires minimales comme *brun*∼*brin* est aujourd'hui bien accomplie à Paris [Malécot et Lindsay, 1976] ; mais leur distinction peut être conservée dans le Sud, en Belgique et en Suisse. Nous avons donc autorisé les variantes [œ] + appendice nasal dans les mots orthographiés avec 'un' ou 'um' : une vingtaine d'items comme *un*, *lundi* ou *parfums*. Nous autorisons ainsi pour le mot *faim* les prononciations [fɛ̃, fɛ̃n, fɛn] tandis que pour le mot *lundi* nous autorisons [lɛ̃di, lɛ̃ndi, lœndi]. En laissant libre les variantes avec [ɛ]/[œ] + appendice nasal pour les /ɛ̃/ et les /œ̃/ sous-jacents, [Boula de Mareüil *et al.*, 2009] ont vérifié que les dénasalisations les plus fréquentes des /ɛ̃/ et /œ̃/ sont bien /ɛ̃/→[ɛn] et /œ̃/→[œn] pour des variétés du nord et du sud de la France.

voyelles nasales
/ɑ̃/→[ɑ̃]∼[ɑ̃n]∼[ɑ̃m]∼[an]∼[am]
/ɛ̃/→[ɛ̃]∼[ɛ̃n]∼[ɛ̃m]∼[ɛn]∼[ɛm]
/œ̃/→[ɛ̃]∼[ɛ̃n]∼[ɛ̃m]∼[œn]∼[œm]
/ɔ̃/→[ɔ̃]∼[ɔ̃n]∼[ɔ̃m]∼[ɔn]∼[ɔm]

TABLEAU 6.1. Variantes de prononciation autorisées pour les voyelles nasales.

Nous avons donc considéré quatre voyelles nasales, pour lesquelles nous avons autorisé des variantes éventuellement dénasalisées, avec un appendice consonantique nasal : les variantes sont données dans le tableau 6.1. Dans tous les cas, les variantes avec [m] sont restreintes à un contexte droit en *p* ou *b*.

6.1.2 Résultats

Le tableau 6.2 montre le pourcentage de voyelles nasales alignées (en lecture et en parole spontanée) comme voyelle nasale, voyelle nasale avec appendice consonantique ou voyelle orale avec appendice consonantique. Deux comportements apparaissent : les locuteurs du sud de la France réalisent davantage d'appendices nasaux, tandis que les autres locuteurs n'en réalisent que très peu. Quand ils prononcent un appendice consonantique nasal, soit dans un peu moins de la moitié des cas, les locuteurs méridionaux réalisent plus souvent une voyelle orale qu'une voyelle nasale. Mis à part dans le Sud, pour lequel les résultats sont comparables quel que soit le style de parole, les locuteurs ont tendance à réaliser plus d'appendices nasaux en parole spontanée qu'en lecture. En lecture, de 90 à 94 % des voyelles nasales sont réalisées de manière canonique en français non-méridional ; en parole spontanée, de 85 à 89 % sont dans ce cas. Rappelons que nous englobons sous l'étiquette « spontané » des entretiens guidés et des conversations libres.

		#occ	%VN	%VN+CN	%VO+CN
lecture	Standard	9 546	92	6	3
	Sud	8 896	**55**	**12**	**33**
	Alsace	2 072	90	4	5
	Belgique	6 500	93	3	4
	Suisse	2 126	94	3	3
spontané	Standard	44 156	85	9	6
	Sud	29 582	**56**	**11**	**33**
	Alsace	5 502	89	4	7
	Belgique	21 504	88	5	7
	Suisse	9 910	88	7	5

TABLEAU 6.2. Pourcentage de voyelles nasales réalisées comme voyelle nasale (VN), voyelle nasale + appendice consonantique nasal (VN+CN), voyelle orale + appendice consonantique nasal (VO+CN) dans le corpus PFC.

La réalisation des voyelles nasales est particulièrement intéressante pour les locuteurs du sud de la France, c'est pourquoi nous avons voulu étudier le comportement propre de chacun des points d'enquête réunis dans la variété Sud du tableau 6.2. La réalisation des voyelles nasales détaillée pour chaque point d'enquête de la variété Sud apparaît dans le tableau 6.3. Nous observons une différence de comportement entre Marseille et Biarritz d'une part, Douzens, Lacaune et Rodez d'autre part. Les deux premiers points d'enquête se comportent plus conformément au standard que les trois derniers. Cette différence pourrait s'expliquer par le substrat dialectal occitan (pour Douzens, Lacaune et Rodez), mais également par une opposition urbain/rural (la grande ville de Marseille face au petit village de Douzens par exemple). Il faut également rappeler que nous n'avons analysé que la lecture du texte pour le point d'enquête de Marseille.

Les résultats de l'alignement du corpus CTS, présentés dans le tableau 6.4, vont dans le même sens : dans le Sud, les locuteurs réalisent plus d'appendices nasaux qu'ailleurs. L'écart entre les variétés est toutefois un peu moins marqué que dans le corpus PFC, comme

	#occ	%VN	%VN+CN	%VO+CN
Biarritz	5 198	72	12	15
Douzens	5 097	**45**	**10**	**44**
Lacaune	3 120	**49**	**13**	**37**
Marseille (lecture)	867	78	10	10
Rodez	4 956	**48**	**12**	**39**

TABLEAU 6.3. Pourcentage de voyelles nasales réalisées comme voyelle nasale (VN), voyelle nasale + appendice consonantique nasal (VN+CN), voyelle orale + appendice consonantique nasal (VO+CN) — détail pour les points d'enquête du Sud dans le corpus PFC, tous styles de parole confondus.

précédemment pour d'autres traits de prononciation : dans le Sud, seules un peu plus de 20 % des occurrences comportent une consonne nasale, et dans ce cas la répartition entre voyelle nasale et orale est équilibrée. Les résultats pour le français standard et l'Alsace sont quasi identiques, avec 90 % de voyelles nasales réalisées sans appendice consonantique.

	#occ	%VN	%VN+CN	%VO+CN
Standard	52 015	90	7	2
Sud	30 192	**77**	**12**	**11**
Alsace	6 225	90	7	3

TABLEAU 6.4. Pourcentage de voyelles nasales réalisées comme voyelle nasale (VN), voyelle nasale + appendice consonantique nasal (VN+CN), voyelle orale + appendice consonantique nasal (VO+CN) dans le corpus CTS.

En conclusion, la réalisation des voyelles nasales semble être un indice robuste permettant de distinguer le sud de la France, où un appendice consonantique nasal est produit plus souvent qu'ailleurs, et ce pour nos deux corpus.

6.2 Réalisation du schwa

Un autre trait caractéristiques des accents du sud de la France est bien décrit dans la littérature : il s'agit de la réalisation de nombreux schwas [Durand *et al.*, 1987]. En français standard, il a été montré que le style informel favorise l'élision du schwa [Boula de Mareüil, 2007] ce que nous pourrons vérifier sur les deux styles de parole du corpus PFC. Dans le système d'alignement que nous avons utilisé, les mêmes modèles acoustiques sont associés au /œ/ et au /ə/ (*cf.* chapitre 2, p. 31).

6.2.1 Variantes proposées

Nous avons autorisé certains schwas à être optionnels dans le dictionnaire de prononciation. Les schwas concernés sont les schwas des monosyllabes (*ce*, *que*...), ceux qui apparaissent en première syllabe de polysyllabes en consonne + schwa (*cheveux*), ceux qui apparaissent en syllabe intérieure de polysyllabe à condition que la règle des trois consonnes soit respectée (le schwa est optionnel par exemple dans *samedi* mais pas dans *vendredi*) et ceux qui apparaissent en fin de polysyllabe. Quand le schwa est maintenu, il peut être prononcé de manière canonique [ə], ou être postériorisé en [ɔ] voir en [o]. Par exemple, pour le mot *relier*, les prononciations [ʁlje, ʁəlje, ʁɔlje, ʁolje] sont autorisées.

Le taux de schwas maintenus obtenu automatiquement a été comparé avec une annotation manuelle réalisée par des linguistes dans le cadre du projet PFC sur un sous-ensemble de nos locuteurs : une corrélation de 0,8 pour le taux de maintien du schwa par locuteur a été obtenue entre les deux méthodes, pour un total de plus de 100 locuteurs [Boula de Mareüil *et al.*, 2009]. Cette bonne corrélation renforce la validité des résultats obtenus en utilisant l'alignement automatique.

6.2.2 Résultats

Les pourcentages de schwas élidés et réalisés comme [ə], [ɔ] ou [o] sont présentés dans le tableau 6.5 pour le corpus PFC. Mise à part la variété Sud, le comportement des locuteurs est assez uniforme : un schwa sur deux est élidé en lecture, et presque trois sur quatre en parole spontanée. Cette observation est conforme à l'hypothèse selon laquelle les schwas tombent plus facilement en parole spontanée. Quand le schwa est réalisé, il est souvent prononcé [ə] (dans un peu plus de 40 % des cas — élision incluse — en lecture, et de 20 % en parole spontanée). Peu d'occurrences sont postériorisées. Les locuteurs du Sud prononcent plus de schwas que les autres. En lecture, le cas le plus fréquent est la réalisation [ə] et non l'élision comme c'est le cas partout ailleurs ; le taux d'élision est pour les deux styles de parole moins élevé que dans les autres variétés de français. Les schwas réalisés par les locuteurs méridionaux sont, en proportion, plus nombreux, quelle que soit la prononciation.

Les résultats détaillés pour les points d'enquête du sud de la France sont présentés dans le tableau 6.6. Les deux groupes qui apparaissent sont un peu différents de ceux qui s'étaient formés pour les voyelles nasales : d'une part les locuteurs de Biarritz et de Rodez, qui élident un peu plus de schwas ; et d'autre part ceux de Douzens et de Lacaune, qui en élident moins. Marseille est à considérer à part, la lecture étant le seul style représenté pour ce point d'enquête : il est normal que les schwas réalisés soient plus nombreux. Mais que ce soit en matière de voyelles nasales ou de schwas, les deux points d'enquête du Languedoc (Douzens et Lacaune) apparaissent les plus conservateurs.

Les données du corpus CTS ont été alignées avec les mêmes variantes de prononciation ; les résultats sont donnés dans le tableau 6.7. Le comportement observé sur le corpus PFC se retrouve sur ces données : deux tiers des schwas tombent pour les locuteurs d'Alsace et de la variété standard, contre seulement moins d'un sur deux pour les locuteurs du Sud. L'écart entre variétés (13 %) est équivalent à celui mesuré sur le corpus PFC en lecture, mais moins important que celui mesuré sur la parole spontanée (environ 20 %), bien que le corpus CTS

		#occ	%élision	%[ə]	%[ɔ]	%[o]
lecture	Standard	6 761	52	41	4	4
	Sud	6 214	**41**	**49**	**6**	**6**
	Alsace	1 560	53	42	2	3
	Belgique	4 844	53	42	3	2
	Suisse	1 635	56	41	3	1
spontané	Standard	20 098	73	21	2	4
	Sud	15 757	**53**	**35**	**5**	**7**
	Alsace	3 241	69	26	3	2
	Belgique	12 247	75	20	2	3
	Suisse	5 644	76	21	2	2

TABLEAU 6.5. Pourcentage de schwas élidés et réalisés comme [ə], [ɔ] ou [o] dans le corpus PFC.

	#occ	%élision	%[ə]	%[ɔ]	%[o]
Biarritz	5 250	57	34	5	5
Douzens	6 577	**45**	**42**	**6**	**6**
Lacaune	3 938	**43**	**43**	**5**	**9**
Marseille (lecture)	1 258	**46**	**44**	**5**	**5**
Rodez	4 944	54	34	4	7

TABLEAU 6.6. Pourcentage de schwas élidés et réalisés comme [ə], [ɔ] ou [o] — détail pour les points d'enquête du Sud dans le corpus PFC, tous styles de parole confondus.

soit constitué de conversations informelles. L'origine moins controlée des locuteurs de ce corpus pourrait expliquer cette différence.

	#occ	%élision	%[ə]	%[ɔ]	%[o]
Standard	66 713	60	33	4	3
Sud	40 924	**47**	**42**	**6**	**6**
Alsace	8 808	60	33	4	3

TABLEAU 6.7. Pourcentage de schwas élidés et réalisés comme [ə], [ɔ] ou [o] dans le corpus CTS.

Comme la réalisation des voyelles nasales, la réalisation du schwa révèle donc un comportement particulier pour les locuteurs de sud de la France, ce sur nos deux corpus. Nous allons la comparer à celle du /ɔ/ dans ce qui suit.

6.3 Réalisation du /ɔ/

Nous avons déjà étudié le timbre du /ɔ/ à travers des mesures de formants dans le chapitre 4. L'étude que nous proposons ici, à travers des variantes d'alignement, aborde la question

d'une autre manière. Nous avons mesuré une antériorisation de ce phonème, notamment pour la variété standard. Les locuteurs du Sud avaient plutôt tendance à rapprocher les timbres de /ɔ/ et /o/. Nous allons examiner si ces phénomènes se confirment également dans le choix des variantes. Cette approche est complémentaire de l'extraction de formants car elle manipule des classes symboliques qui sont intéressantes pour des interprétations catégorielles en phonologie.

6.3.1 Variantes proposées

Nous avons autorisé dans le dictionnaire de prononciation trois variantes : une réalisation canonique [ɔ], une variante antériorisée [œ] ainsi qu'une prononciation [o] qui pourra rendre compte de la neutralisation de l'opposition phonologique /o/∼/ɔ/ dans le Sud [Durand et Lyche, 2004]. Pour le mot *sol*, par exemple, les réalisations suivantes sont autorisées : [sɔl, sœl, sol].

6.3.2 Résultats

D'après les résultats présentés dans le tableau 6.8 pour le corpus PFC, la prononciation canonique [ɔ] est la plus souvent choisie pour les locuteurs de la variété standard, de Belgique et de Suisse. Cette réalisation est un peu plus fréquente en lecture qu'en parole spontanée où la prononciation est moins surveillée. Ce sont les Suisses qui réalisent la plus forte proportion de [ɔ] : $3/4$ des occurrences en lecture et plus de la moitié en parole spontanée. Quand ils n'adoptent pas cette prononciation, ils ont tendance à antérioriser en prononçant [œ]. Les Belges antériorisent le /ɔ/ en lecture et vont vers le [o] en spontané. Les locuteurs standard réalisent à peu près autant de [o] et de [œ] en lecture, et antériorisent davantage en parole spontanée, ce qui est cohérent avec les mesures de formants.

		#occ	%[ɔ]	%[œ]	%[o]
lecture	Standard	1 339	55	21	24
	Sud	1 355	**35**	**5**	**61**
	Alsace	319	**28**	**10**	**62**
	Belgique	986	50	20	7
	Suisse	323	73	20	7
spontané	Standard	4 756	40	34	26
	Sud	3 698	**39**	**13**	**49**
	Alsace	749	**38**	**21**	**41**
	Belgique	2 617	37	21	33
	Suisse	1 326	55	31	14

TABLEAU 6.8. Pourcentage de /ɔ/ alignés comme [ɔ], [œ] ou [o] pour les locuteurs du corpus PFC.

En Alsace et dans le Sud, la prononciation la plus fréquente est [o], que ce soit en lecture ou en parole spontanée. Ce n'est pas surprenant pour les locuteurs du Sud, qui ont tendance à neutraliser l'opposition phonologique /o/∼/ɔ/ ; ce phénomène apparaissait également sur les

triangles vocaliques. Les réalisations des Alsaciens sont difficiles à mettre en relation avec les mesures de formants, étant donné que ces derniers n'étaient pas aisément comparables entre l'Alsace et les autres points d'enquête (*cf.* chapitre 4). Pour autant, la prononciation [o] n'est pas aberrante en Alsace : [Walter, 1982] mentionne des opposition /ɔ/∼/o/ fluctuantes chez certains de ses locuteurs alsaciens.

Comme l'Alsace ne constitue qu'un seul point d'enquête, nous ne pouvons pas détailler les résultats de cette variété. En revanche, nous pouvons le faire pour le Sud. Ces résultats détaillés sont présentés dans le tableau 6.9 : nous remarquons que les locuteurs de Lacaune réalisent une forte proportion de /ɔ/ comme [o]. Viennent ensuite Marseille et Rodez, et enfin, les locuteurs de Biarritz et Douzens, qui réalisent le plus de [ɔ] canoniques. La réalisation [o] reste la plus fréquente pour les cinq points d'enquête méridionaux.

	#occ	%[ɔ]	%[œ]	%[o]
Biarritz	1 135	45	9	47
Douzens	1 498	43	11	46
Lacaune	927	**24**	**11**	**65**
Marseille (lecture)	283	36	6	58
Rodez	1 209	35	12	53

TABLEAU 6.9. Pourcentage de /ɔ/ alignés comme [ɔ], [œ] ou [o] — détail pour les points d'enquête du Sud dans le corpus PFC.

Les mêmes variantes ont été appliquées au corpus CTS (tableau 6.10). Les réalisations observées pour les trois variétés présentes dans le corpus sont très proches. Si le taux de [ɔ] est effectivement plus bas en Alsace et dans le sud, et le taux de [o] plus élevé pour ces mêmes régions, les différences ne sont que de quelques pourcents. Sans aller dans le sens inverse des tendances du corpus PFC, ces résultats ne les renforcent que peu.

	#occ	%[ɔ]	%[œ]	%[o]
Standard	13 066	75	9	14
Sud	8 151	71	7	22
Alsace	1 826	74	7	19

TABLEAU 6.10. Pourcentage de /ɔ/ alignés comme [ɔ], [œ] ou [o] pour les locuteurs du corpus CTS.

Les locuteurs du sud de la France et d'Alsace du corpus PFC ont tendance à préférer la variante [o], tandis que comparativement on observe davantage de [œ] en français standard. Cette tendance ne se retrouve que peu dans le corpus CTS. Pourtant, les triangles vocaliques montraient que les locuteurs des variétés Sud et Alsace antériorisaient moins le /ɔ/ que ceux du français standard (voir chapitre 4). Dans le cas présent, ce sont les mesures acoustiques qui mettent en avant des différences plus marquées entre variétés.

6.4 Réalisation des consonnes occlusives et fricatives

Nous avons également étudié la réalisation des consonnes sourdes (occlusives /p t k/ ou fricatives/f s ʃ/) et de leurs équivalents sonores (occlusives /b d g/ ou fricatives /v z ʒ/). Cette étude est complémentaire à celle que nous avons commencée au chapitre 5 avec les taux de voisement. Nous n'avons pas inclus dans cette section la réalisation du /ʁ/ qui pose des questions un peu différentes et que nous examinerons dans la section suivante.

Les consonnes sourdes et sonores ont été étudiées dans deux études dont le but était une analyse du phénomène d'assimilation en français. Des variantes similaires à celles que nous venons de décrire ont été proposées par [Adda-Decker et Hallé, 2007], qui ont calculé le taux de consonnes ayant subi une modification entre la forme de surface et la forme sous-jacente, sur un corpus de français standard. Les auteurs ont de même étudié ce phénomène à travers des taux de voisement [Hallé et Adda-Decker, 2007].

6.4.1 Variantes proposées

Pour chacune des consonnes occlusives et fricatives citées ci-dessus, nous avons proposé comme variante la contrepartie sourde ou sonore : la liste complète est donnée dans le tableau 6.11.

occlusives		fricatives	
sourdes	sonores	sourdes	sonores
/p/→[p]∼[b]	/b/→[p]∼[b]	/f/→[f]∼[v]	/v/→[f]∼[v]
/t/→[t]∼[d]	/d/→[t]∼[d]	/s/→[s]∼[z]	/z/→[s]∼[z]
/k/→[k]∼[g]	/g/→[k]∼[g]	/ʃ/→[ʃ]∼[ʒ]	/ʒ/→[ʃ]∼[ʒ]

TABLEAU 6.11. Liste des variantes proposées pour l'étude des consonnes occlusives et fricatives.

6.4.2 Résultats

Le pourcentage de consonnes sourdes alignées comme sonores, et inversement, est donné dans le tableau 6.12, pour les occlusives et les fricatives. Comme nous l'avons déjà constaté pour d'autres variantes, le pourcentage de réalisation non canoniques est plus important en parole spontanée, moins contrôlée que la lecture. La réalisation des consonnes sourdes ne montre pas de différence selon les variétés, quel que soit le style de parole, cependant le taux de consonnes sourdes alignées comme sonores est plus élevé pour les fricatives que pour les occlusives. Des différences apparaissent en revanche pour les consonnes sonores : les Alsaciens présentent une proportion importante de fricatives et plus encore d'occlusives sonores alignées comme sourdes.

Dans sa thèse, [Vieru-Dimulescu, 2008] présente également des résultats concernant l'alignement de consonnes sonores comme sourdes. Pour les locuteurs du français standard, les taux d'occlusives sonores alignées comme sourdes sont cohérent avec nos résultats. Pour

		#sourdes	%sourdes ↓ sonores		#sonores	%sonores ↓ sourdes	
			occl.	fric.		occl.	fric.
lecture	Standard	10 925	9	14	4 890	8	10
	Sud	11 024	7	9	4 908	7	9
	Alsace	2 430	4	8	1090	**40**	**23**
	Belgique	10 912	4	9	5 035	13	12
	Suisse	2 732	3	5	1 197	8	12
spontané	Standard	37 757	14	18	15 766	14	12
	Sud	27 597	16	16	12 008	12	11
	Alsace	4 474	10	20	1 773	**46**	**24**
	Belgique	30 410	11	19	12 301	21	15
	Suisse	10 166	8	12	4 570	12	13

TABLEAU 6.12. Pourcentage de consonnes sonores alignées comme sourdes et de consonnes sourdes alignées comme sonores pour les occlusives et les fricatives du corpus PFC.

les fricatives, nos pourcentages sont un peu plus bas. Les mêmes chiffres sont disponibles pour des locuteurs allemands parlant français, connus pour dévoiser les consonnes sonores. Aussi bien pour les occlusives que pour les fricatives, nos locuteurs alsaciens entrent dans la fourchette de taux indiquée pour les Allemands. Il est intéressant de noter que le taux de réalisations dévoisées des Alsaciens atteint les mêmes proportions que pour les locuteurs allemands, en matière d'occlusives comme de fricatives.

	#sourdes	%sourdes ↓ sonores		#sonores	%sonores ↓ sourdes	
		occlusives	fricatives		occlusives	fricatives
Standard	130 796	9	9	58 866	9	10
Sud	79 604	8	7	35 581	9	10
Alsace	17 281	8	9	7 515	**14**	**12**

TABLEAU 6.13. Pourcentage de consonnes sonores alignées comme sourdes et de consonnes sourdes alignées comme sonores pour les occlusives et les fricatives du corpus CTS.

Les mêmes mesures, effectuées sur le corpus CTS, ne révèlent pas de grandes différences selon les variétés (voir tableau 6.13). Les Alsaciens ont un pourcentage de consonnes sonores alignées comme sourdes plus élevé que les autres locuteurs, mais l'écart ne dépasse pas 5 %.

Nous n'aboutissons pas aux mêmes conclusions avec cette approche qu'avec les taux de voisement calculés au chapitre 5. Les résultats présentés ici pour le corpus PFC sont plus en accord avec l'hypothèse selon laquelle les Alsaciens dévoiseraient les occlusives sonores

[Walter, 1982]. Nous n'avions pas pu retrouver cette tendance de manière aussi nette dans les taux de voisement. Ces variantes, appliquées au corpus CTS, donnent, comme pour les voyelles, des tendances moins marquées que sur le corpus PFC.

6.5 Réalisation du /ʁ/

Le /ʁ/ est un phonème très saillant perceptivement et fréquent en français, mais il est difficile à caractériser du point de vue phonétique et phonologique. Dans le chapitre 5, nous avons étudié le /ʁ/ en calculant des taux de voisement. Nous étudions ici ce phonème à l'aide de variantes de prononciation, comme nous l'avons fait pour les consonnes occlusives et fricatives dans la section précédente. Contrairement à ces dernières, pour lesquelles nous avons pu proposer des alternatives voisées ou non-voisées appartenant à l'inventaire des phonèmes du français, nous n'avons pas de prononciation alternative du /ʁ/ dans le système français.

6.5.1 Variantes proposées

Pour permettre l'étude de la réalisation du /ʁ/, nous faisons appel à des xénophones — phonèmes issus du système phonémique d'une autre langue. Nous avons à notre disposition de modèles acoustiques étrangers, entraînés sur des quantités de données importantes. Nous avons choisi de proposer la jota [x] et le 'ɾ' battu [ɾ], tous deux issus du modèle acoustique espagnol, comme réalisations alternatives au [ʁ] du modèle acoustique français. Ainsi, pour le mot *car*, les variantes [kaʁ, kaɾ, kax] sont autorisées.

6.5.2 Résultats

Les réalisations du /ʁ/ sont indiquées dans le tableau 6.14. Quelle que soit la variété considérée, la variante française [ʁ] est choisie dans la majorité des cas. La proportion de réalisations canoniques est légèrement moins élevée en parole spontanée qu'en lecture, tendance que nous avons déjà pu relever de manière plus marquée pour d'autres phonèmes. Les locuteurs qui réalisent le plus de [ʁ] sont les Suisses, suivis de ceux du français standard. Ceux qui en réalisent le moins sont les Alsaciens, dans les deux styles. Lorsqu'ils réalisent un /ʁ/ non-standard, les locuteurs du Sud montrent plutôt un [ɾ]. Les Belges ont des prononciations plus proches du modèle de la jota espagnole. Dans le chapitre précédent, nous avons remarqué un taux de voisement du /ʁ/ plus bas pour les locuteurs belges en parole spontanée. Cette observation est en accord avec la proportion de [x] plus élevée pour ces locuteurs. Les Alsaciens montrent une proportion proche pour les deux réalisations, qui s'entendent effectivement bien dans nos données.

Nous avons constaté que les points d'enquête belges se comportent tous les trois de manière similaire et, de ce fait, nous n'avons pas détaillé leurs résultats ici. Tel n'est pas le cas des points d'enquête du Sud, pour lesquels les résultats détaillés sont présentés dans le tableau 6.15. Les points d'enquête de Marseille et de Rodez ont un comportement plutôt proche du standard, avec beaucoup de réalisations canoniques. Au contraire, les locuteurs

		#occ	%[ʁ]	%[ɾ]	%[x]
lecture	Standard	5 693	80	11	8
	Sud	4 782	76	14	10
	Alsace	1 170	75	12	12
	Belgique	3 460	78	10	13
	Suisse	1 250	83	6	11
spontané	Standard	16 790	76	13	11
	Sud	9 808	71	17	13
	Alsace	1 380	63	20	19
	Belgique	3 242	65	14	21
	Suisse	4 566	78	9	13

TABLEAU 6.14. Pourcentage de /ʁ/ réalisés [ʁ], [ɾ] ou [x] dans le corpus PFC.

de Douzens réalisent 30 % de [ɾ], et seulement 60 % de [ʁ]. Il est également intéressant de noter que pour les locuteurs de Biarritz, la réalisation non-standard la plus fréquente n'est pas le [ɾ], mais le [x]. Ces deux derniers phénomènes sont bien audibles : plusieurs locuteurs de Douzens « roulent » les 'r', tandis qu'on entend plus de 'r' proches d'une jota chez les locuteurs de Biarritz.

	#occ	%[ʁ]	%[ɾ]	%[x]
Biarritz	4 962	75	11	14
Douzens	5 316	58	29	13
Lacaune	3 510	69	17	14
Marseille (lecture)	1 086	87	8	5
Rodez	4 994	82	8	10

TABLEAU 6.15. Pourcentage de /ʁ/ réalisés [ʁ], [ɾ] ou [x] pour les points d'enquête du Sud dans le corpus PFC.

Nous avons identifié perceptivement neuf locuteurs roulant les 'r' dans le corpus PFC : un locuteur de Boersch, un de Gembloux, un de Biarritz, deux de Lacaune et quatre de Douzens, tous assez âgés. Nous avons examiné les résultats obtenus pour l'alignement des variantes du /ʁ/ pour ces neuf locuteurs.

Pour tous, la variante qui est la plus souvent alignée est [ɾ], en plus ou moins forte proportion (de 40 à 77 %). Viennent ensuite le [ʁ] puis le [x]. Pour les autres locuteurs du corpus, la variante [ɾ] n'a été alignée que dans 30 % des cas au maximum. Ces résultats sont encourageants car, perceptivement, il n'est pas facile de caractériser avec précision les différentes réalisations du /ʁ/ : la variante [ɾ] semble bien être choisie globalement par le système pour les locuteurs pour lesquels nous l'attendions, et la jota n'a pas été sélectionnée à sa place.

Pour l'Alsace, les deux prononciations ([ʁ] et [x]) sont audibles dans le corpus CTS . Nous n'avons cependant pas fait les mêmes alignements, qui sont surtout intéressants pour la Belgique, non couverte par notre corpus téléphonique.

6.6 Conclusion

Grâce à l'utilisation de variantes de prononciation, nous avons pu quantifier certains phénomènes bien connus : la dénasalisation des voyelles nasales, le maintien du schwa ou encore la neutralisation de l'opposition /o/∼/ɔ/ dans le sud de la France, ainsi que le dévoisement des consonnes fricatives et surtout occlusives sonores en Alsace. Nous avons également obtenu des résultats intéressants concernant l'antériorisation du /ɔ/, phénomène moins documenté. Nous avons constaté dans l'ensemble que les réalisation canoniques sont plus nombreuses en lecture qu'en parole spontanée, qui est une façon plus naturelle de parler.

Un certain nombre de variantes que nous avons étudiées permettent de caractériser les variétés du sud de la France. En regardant les résultats point d'enquête par point d'enquête, nous avons pu constater des différences de comportement. Cependant, ces différences ne paraissent pas forcément suffisantes pour permettre d'identifier tel ou tel point. Perceptivement, nos auditeurs n'y étaient pas bien parvenus (*cf.* chapitre 4). Les différentes variables ne permettent pas non plus de détacher les mêmes régions méridionales des autres. Nous ne chercherons plus par la suite à les discriminer.

L'étude de certaines variantes de prononciation était complémentaire d'études précédentes à base de mesures acoustiques. Dans certains cas, les deux approches donnent des résultats convergents : par exemple, les réalisations du /ɔ/ obtenues avec les variantes de prononciation correspondent aux résultats obtenus avec les formants. Dans d'autre cas, nous n'aboutissons pas aux mêmes conclusions selon l'approche adoptée. Ainsi, l'étude des consonnes à partir des variantes d'alignement met en valeur un phénomène (le dévoisement des consonnes sonores) pour l'Alsace, tandis que l'étude fondée sur la fréquence fondamentale ne permettait pas de distinction aussi nette de cette région. Dans ce cas précis, c'est l'approche à base d'alignement qui corrobore l'hypothèse linguistique.

Un même phénomème linguistique peut donc être mesuré et caractérisé de plusieurs manières. Dans le contexte d'une tâche de classification, certains indices sont-ils plus efficaces que d'autres pour discriminer les variétés régionales ? Quelle approche permet d'obtenir les indices les plus pertinents ? Nous essaierons de répondre à ces questions dans le chapitre suivant.

Chapitre 7

Classification

Nous avons, dans les chapitres précédents, mis en évidence un certain nombre d'indices montrant des différences entre variétés régionales de français. Ces indices ont jusqu'ici été présentés pour chacune des variétés de français que nous étudions (ou éventuellement certains points d'enquête particuliers du corpus PFC). En vue d'une tâche de classification, ces paramètres, calculés pour chacun de nos locuteurs, pourraient se montrer trop variables pour permettre une identification automatique. De ce fait, nous pouvons nous demander si les indices que nous avons calculés sont suffisants pour classifier nos locuteurs et ce avec quelle précision. La réponse à cette question dépend des conditions dans lesquelles nous nous plaçons : nous pouvons essayer de retrouver l'origine d'un locuteur parmi trois possibilités, parmi cinq, ou parmi plus encore…

Dans ce chapitre, nous allons utiliser des techniques de classification automatique. Nous donnerons en entrée un certain nombre d'indices, calculés pour chaque locuteur. Se montreront-ils suffisants pour une tâche de classification ? Quels indices seront les plus discriminants ? Faut-il en privilégier certains ? Arriverons-nous à des taux d'identification corrects à partir de quelques indices motivés linguistiquement ? Pourrons-nous nous rapprocher des résultats de la perception humaine ? Bien que les conditions ne soient pas les mêmes, il sera intéressant de comparer les taux d'identification automatique avec ceux du chapitre 3.

Nous présenterons tout d'abord les méthodes utilisées pour classifier nos données, avant d'étudier les résultats sur nos corpus.

7.1 Méthode

Étant donnée une instance (*i.e.* un locuteur) et plusieurs classes (*i.e.* nos 5 variétés de français dans le cas où nous voulons identifier l'origine parmi 5 possibilités), nous cherchons à savoir quelle classe attribuer à cette instance. Pour ce faire, nous avons choisi deux classifieurs adaptés à cette tâche : les arbres de décision et les SVM. Nous présentons brièvement leur fonctionnement dans la section 7.1.1. Nous décrirons ensuite la méthode de validation adoptée dans la section 7.1.2.

7.1.1 Classifieurs

Comme précédemment pour nos analyses statistiques, nous avons travaillé avec le logiciel R [R Development Core Team, 2008] qui propose des librairies implémentant entre autres les deux classifieurs qui nous intéressent.

Arbres de décision

Les arbres de décision sont des classifieurs qui ont l'avantage de pouvoir être présentés sous une forme graphique aisément interprétable par un humain. Un arbre est construit à partir d'un ensemble d'apprentissage ; une fois mis au point, il peut servir à prédire la classe de nouvelles observations. Les arbres sont de la forme suivante : chaque nœud correspond à un test, comparant la valeur d'un attribut à un seuil constant. Les branches issues de ce nœud correspondent à la valeur booléenne du test, vrai ou faux. Chaque feuille de l'arbre est associée à une classe. La croissance de l'arbre est arrêtée dès qu'il n'y a plus de partition possible ou que le nombre d'observations dans une feuille est inférieur à une valeur seuil. Pour classifier une nouvelle instance grâce à l'arbre construit, il suffit d'appliquer chaque test aux valeurs de ses attributs. La classe attribuée à l'instance sera celle de la feuille qu'elle aura atteinte.

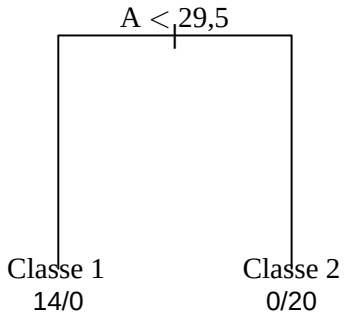


FIGURE 7.1. Exemple d'arbre de décision sous forme graphique.

La figure 7.1 représente graphiquement un exemple simple d'arbre. Les données sont réparties en deux classes, nommées Classe 1 et Classe 2. L'arbre n'a qu'un nœud, indiquant le test $A < 29{,}5$: si l'attribut A est inférieur à 29,5, le test est vérifié et l'instance est envoyée dans la branche gauche ; dans le cas contraire, l'instance part dans la branche droite. Les arbres présentés ci-après sont évidemment moins simples, mais à chaque fois l'ordre est le même : branche gauche si le test est vérifié, branche droite sinon. L'arbre rend parfaitement compte des données qui ont servi à le construire : la branche gauche comprend 14 instances de la Classe 1 et 0 de la Classe 2. La branche droite comprend 20 instances de la Classe 2 et 0 de la Classe 1. Ces chiffres sont indiqués sous chaque feuille par des nombres donnés dans l'ordre alphabétique des classes.

Pour construire des arbres de décision, nous avons utilisé la fonction `rpart` de la librairie du même nom du logiciel R [R Development Core Team, 2008]. Cette fonction renvoie un

arbre construit d'après l'algorithme CART (*Classification And Regression Tree*) [Breiman *et al.*, 1984].

SVM (Séparateurs à Vaste Marge)

Les Séparateurs à Vaste Marge, également appelés Machines à vecteurs de support (de l'anglais *Support Vector Machines*), peuvent entre autre être utilisés pour des tâches de classification supervisée, comme celle qui nous intéresse ici.

Ces classifieurs fonctionnent de la manière suivante : l'algorithme trouve un hyperplan permettant de séparer au mieux les données d'apprentissage en fonction de leur classe. Les frontières de séparation sont choisies de façon à maximiser la marge, c'est-à-dire la distance entre la frontière de séparation et les échantillons les plus proches, également appelés *support vectors*. Dans le cas où les données traitées ne sont pas séparable linéairement, il est possible d'utiliser la fonction noyau (*kernel*) du SVM pour les projeter dans un espace de dimension plus grande, dans lequel existera un séparateur linéaire.

À la base, les SVM sont conçus pour fonctionner avec des données réparties en deux classes. Ceux que nous utilisons ont été adaptés afin de pouvoir fonctionner en multi-classe. La méthode utilisée est celle du un contre un (*one-against-one*, [Hsu et Lin, 2002]) : si les données sont réparties en k classes, $k(k-1)/2$ classifieurs binaires sont entraînés, et un algorithme de vote choisit la classe finalement attribuée à l'observation.

Pour mettre en œuvre ces classifieurs, nous avons utilisé la fonction `svm` de la librairie `e1071` du logiciel R, qui est une interface de la librairie `libsvm` [Chang et Lin, 2001]. Nous avons de plus choisi une fonction noyau polynomiale.

7.1.2 Validation croisée et *leave-one-out*

Dans une tâche de classification en bonne et due forme, les données doivent être réparties en deux ensembles : l'un sert à l'apprentissage du classifieur, l'autre au test. Cette répartition est difficilement applicable à nos données, c'est pourquoi nous avons utilisé l'ensemble de nos enregistrements dans les chapitres précédents, sans en laisser de côté pour le test. En effet, si nous nous en tenons à la classification de cinq variétés de français, nous n'avons que peu d'observations pour certaines classes (l'Alsace et la Suisse ne comptent chacune qu'une douzaine de locuteurs), et de ce fait le classifieur ne peut être correctement entraîné.

La validation croisée, ou *cross-validation*, est une méthode permettant de remédier à ce problème. Le principe en est le suivant : l'ensemble des données est divisé en N sous-ensembles. Pour chacun d'entre eux, la même procédure est appliquée : le sous-ensemble concerné est utilisé pour le test, tandis que les $N-1$ autres servent à l'apprentissage. Il est possible de déterminer un score d'identification en calculant la moyenne des N scores obtenus.

Le cas particulier de validation croisée dans lequel les sous-ensembles considérés sont réduits à une seule observation (dans notre cas, un seul locuteur) est appelé *leave-one-out*. C'est cette méthode que nous avons choisi d'appliquer dans nos travaux.

7.2 Classification des locuteurs

7.2.1 Paramètres donnés en entrée

Dans les chapitres précédents, nous avons étudié de nombreux paramètres pour les différentes variétés de français auxquelles nous nous sommes intéressée. Pour chacun de nos locuteurs, nous avons proposé une ou plusieurs mesures reflétant divers traits de prononciation. Certains paramètres posent un problème de fiabilité : les différences que nous avons observées au chapitre 5 concernant les taux de voisements sont-elles dues à des conditions d'enregistrement variables en qualité ou à l'accent des locuteurs ? D'autres, comme les formants de certaines voyelles, ne semblent pas à première vue utiles pour identifier l'origine des locuteurs. D'autres encore, comme les taux de différences locales (pour F_0, l'intensité et la durée) liés à la prosodie et décrits en détails au chapitre 5, ont nécessité de fixer des seuils pour ne pas multiplier les redondances. Étant données ces observations, nous proposons un ensemble d'attributs étendu, comprenant toutes les mesures auxquelles nous nous sommes intéressée, et un ensemble restreint, dans lequel nous avons uniquement conservé certaines mesures au vu des résultats des chapitres précédents.

Tous les attributs que nous avons pris en considération sont indiqués ci-dessous. Les attributs marqués d'une étoile (*) font uniquement partie de l'ensemble étendu ; les autres appartiennent aux deux ensembles.

Valeurs des formants (20 attributs) : (*) la valeur moyenne des deux premiers formants des 10 voyelles orales du français. Seul le second formant du /ɔ/ est conservé dans l'ensemble restreint.

Voisement des consonnes (3 attributs) : le taux de voisement des consonnes suivantes :

- (*) consonnes sourdes /p t k f s ʃ/ ;
- (*) consonnes sonores /b d g v z ʒ/ ;
- (*) /ʁ/.

Paramètres liés à la prosodie (7 attributs) :
- (*) la durée moyenne des phonèmes ;
- la durée moyenne des attaques des disyllabes ou plus précédés d'un clitique ;
- le taux de ΔF_0 positif pour les disyllabes ou plus précédés d'un clitique ;
- le taux de $\Delta\text{durée}$ positif pour les trisyllabes ou plus précédés d'un clitique ;
- le taux de $\Delta\text{intensité}$ positif pour les disyllabes ou plus précédés d'un clitique ;
- le taux de $\Delta\text{durée}_f$ positif pour les trisyllabes ou plus précédant une pause ;
- le taux de $\Delta\text{durée}_p$ positif pour les trisyllabes ou plus précédant une pause.

Variantes de prononciation (8 attributs) : les taux de

- voyelles nasales réalisées avec un appendice consonantique nasal ;
- schwas élidés ;
- /ɔ/→[o], /ɔ/→[œ] ;
- /p t k f s ʃ/→[b d g v z ʒ], /b d g v z ʒ/→[p t k f s ʃ] ;
- /ʁ/→[ɾ] et /ʁ/→[x] [1].

1. Attributs uniquement calculés sur le corpus PFC (voir section 6.5).

Au total, nous avons un ensemble restreint de 15 attributs et un ensemble étendu de 38 attributs pour représenter chaque locuteur[2]. Les valeurs des attributs sont calculées sur la quantité de parole disponible pour chaque locuteur, soit en moyenne 3 minutes de lecture et 13 minutes de parole spontanée (ou 16 minutes tous styles de parole confondus) pour les locuteurs du corpus PFC, et 7 minutes de parole spontanée pour les locuteurs du corpus CTS. Le nombre de locuteurs de nos deux corpus est indiqué dans les tableaux 2.1, p. 29 et 2.2, p. 30.

7.2.2 Classification *leave-one-out* sur le corpus PFC

Le corpus PFC se prête bien à la méthode de validation croisée *leave-one-out* qui permet de prendre en compte des classes comportant peu d'instances. Pour ce corpus, nous avons examiné deux cas de figure : la classification en cinq puis en trois variétés de français.

Classification en 5 variétés

Dans un premier temps, nous avons voulu classifier nos données en 5 variétés de français : Standard, Alsace, Belgique, Suisse et Sud. Nous avons considéré les ensembles restreints et étendus d'attributs, et nous avons pris en compte différents styles de parole pour le calcul des attributs : ils peuvent être calculés sur la parole lue, spontanée, ou encore sur la totalité de la parole disponible. Le taux d'instances correctement classifiées par nos deux classifieurs, arbres de décision et SVM, est donné dans le tableau 7.1 pour différents styles de parole.

	ensemble d'attributs	lecture ($\sim$ 3 min)	spontané ($\sim$ 13 min)	tout ($\sim$ 16 min)
Arbres	restreint	56	60	59
	étendu	69	73	67
SVM	restreint	64	74	78
	étendu	70	75	82

TABLEAU 7.1. Pourcentage d'instances correctement classifiées en 5 variétés pour la lecture, la parole spontanée et ces deux styles confondus, par les arbres de décision et les SVM, avec les ensembles d'attributs restreints et étendus. Les pourcentages sont donnés par rapport à un peu moins de 170 locuteurs. Entre parenthèses, la quantité de parole disponible en moyenne pour chaque locuteur.

Dans l'ensemble, les taux d'identification sont plutôt satisfaisants (le taux de hasard est de 20 % pour une décision entre cinq classes équiprobables). Ils sont tous plus élevés que ceux que nous avons pu mesurer en perception au chapitre 3, bien que la comparaison soit à prendre avec précaution. En effet, nos auditeurs ne disposaient que d'une dizaine de secondes de parole et n'étaient pas entraînés pour la tâche qu'ils avaient à accomplir.

2. Certains attributs n'ont pu être calculés pour quelques uns de nos locuteurs pour lesquels nous disposions de trop peu de données. Nous avons omis ces rares locuteurs lors de la classification.

Au contraire, nos classifieurs ont bénéficié du *leave-one-out*, qui leur donne des conditions favorables en maximisant la taille de l'ensemble d'apprentissage.

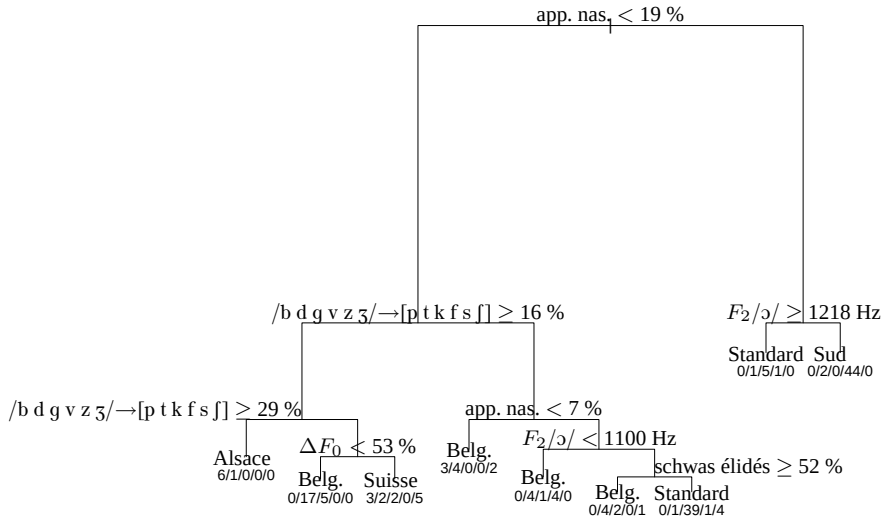
Pour des configurations identiques, les SVM donnent de meilleurs résultats que les arbres de décision (ils nécessitent toutefois un temps de calcul plus important, notamment pour la phase d'apprentissage). Précisons ici que les classifieurs utilisés n'ont pas été finement réglés pour obtenir les meilleurs résultats : nous avons conservé une grande partie des réglages par défaut. Les résultats semblent meilleurs lorsque l'ensemble étendu d'attributs est pris en compte (le fait d'ajouter des attributs dans un classifieur n'entraîne pas forcément une amélioration de ses performances).

Pour les deux classifieurs, les taux d'identification les plus faibles sont observés pour la lecture. En effet, la quantité de parole pour ce style est limitée et les valeurs des attributs sont de ce fait calculées sur un nombre d'occurences moindre. Les SVM montrent de meilleures performances (jusqu'à 82 % d'identification) sur les deux styles de parole confondus, configuration dans laquelle les attributs sont calculés sur la plus grande quantité de données. Pour les arbres de décision, les meilleurs taux d'identification (jusqu'à 73 %) s'observent en parole spontanée, tandis que la combinaison des différents styles de parole ne permet pas d'augmenter les taux.

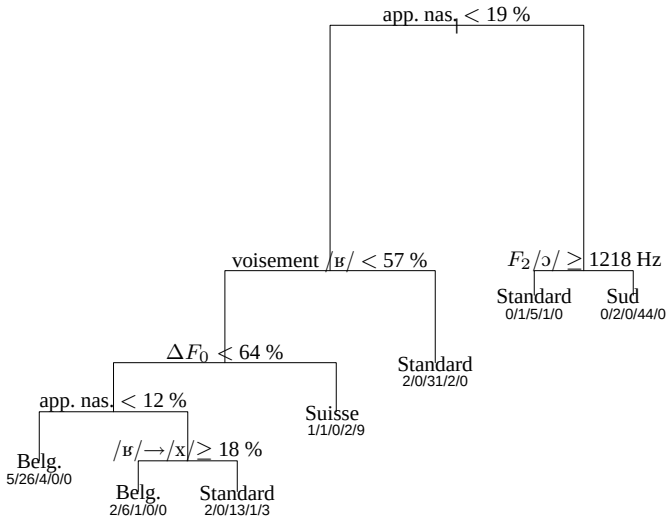
Bien que les arbres de décision n'aient pas montré les meilleures performances sur nos données, nous les avons retenu pour des raisons « pédagogiques » : les noeuds présentent les indices utilisés et la hauteur des branches indique une plus ou moins grande différence entre les feuilles. La figure 7.2 montre les arbres ainsi construits à partir des ensembles d'attributs restreint et étendu, pour tous les locuteurs du corpus PFC. Sous chaque feuille est indiqué le nombre de locuteurs de chacune des 5 classes inclus dans cette feuille. Les chiffres sont donnés dans l'ordre alphabétique (Alsace, Belgique, Standard, Sud, Suisse). Dans les deux représentations, les locuteurs du Sud se distinguent d'abord grâce au taux de réalisation d'appendices nasaux puis grâce au second formant du /ɔ/.

Les traits utilisés pour caractériser les autres locuteurs diffèrent selon l'ensemble d'attributs considéré. Dans le cas de l'ensemble restreint (figure 7.2(a)), les Alsaciens, les Suisses et une partie des Belges se démarquent par le dévoisement des consonnes sonores. Les Alsaciens sont ensuite isolés par un dévoisement encore plus important ; les Suisses par l'accentuation initiale (montée de F_0). Les traits cités jusqu'à présent font sens d'après les interprétations linguistiques que nous avons pu faire dans les chapitres précédents. Enfin, une partie des Belges est classée de manière relativement proche du standard. Les traits qui différencient ces locuteurs belges ne sont pas aussi pertinents que les autres du point de vue linguistique. En revanche, le fait que ces locuteurs soient proches du standard se comprend assez bien au regard de nombre de résultats en analyses acoustiques et en perception. Il reste encore à mettre nos résultats de classification en relation avec le degré d'accent des locuteurs.

L'arbre construit avec l'ensemble étendu (représenté figure 7.2(b)) illustre l'utilisation du taux de voisement du /ʁ/ pour différencier les variétés de l'Est. Le doute demeure : la variation de taux de voisement n'est-elle pas due aux conditions d'enregistrement plutôt qu'à l'accent des locuteurs ? Ce paramètre sera-t-il robuste lors de l'ajout de nouveaux points d'enquête ? Si les locuteurs suisses sont encore identifiés par un corrélat mélodique de l'accentuation initiale, les Alsaciens ne s'en sont pas suffisamment démarqués pour générer une



(a) ensemble d'attributs restreint



(b) ensemble d'attributs étendu

FIGURE 7.2. Arbres de décision construits à partir de tous les locuteurs du corpus PFC pour une classification en 5 variétés.

feuille dans l'arbre.

Ces représentations graphiques nous montrent que certaines variétés sont plus proches que d'autres au niveau de la structure des arbres. Cette observation invite à une comparaison avec les tests de perception. De même que nous avons étudié les confusions des auditeurs humains, quelles sont celles des classifieurs automatiques ? Les matrices de confusion, présentées dans le tableau 7.2 pour les arbres de décision et les SVM, permettent de faire le parallèle entre les résultats en identification par les humains et par la machine. Les résultats sont donnés en pourcentages bien que les nombres de locuteurs soient inégaux d'une variété à l'autre.

	origine \ réponse	Arbres					SVM				
		Standard	Alsace	Belgique	Suisse	Sud	Standard	Alsace	Belgique	Suisse	Sud
restreint	Standard	**69**	0	15	7	9	**78**	2	13	6	2
	Alsace	**38**	15	**38**	8	0	8	**67**	17	0	8
	Belgique	**42**	3	33	6	17	22	8	**61**	0	8
	Suisse	**38**	0	25	33	0	25	0	8	**67**	0
	Sud	8	0	8	0	**84**	2	0	2	0	**96**
étendu	Standard	**80**	0	11	0	9	**81**	7	6	0	6
	Alsace	33	0	**58**	8	0	8	**58**	25	8	0
	Belgique	17	14	**61**	3	6	11	6	**69**	6	8
	Suisse	15	0	8	**75**	0	25	0	0	**75**	0
	Sud	16	0	6	6	**72**	2	0	0	0	**98**

TABLEAU 7.2. Matrice de confusion (%) obtenue par validation croisée *leave-one-out* sur l'ensemble des données PFC classifiées en 5 variétés, en considérant les ensembles restreint et étendu d'attributs, avec les arbres de décision et les SVM.

Quel que soit l'ensemble d'attributs considéré, la réponse majoritairement donnée par le SVM est toujours la bonne. Pour ce classifieur, l'origine des locuteurs est correctement identifiée dans au moins 58 % des cas ; les locuteurs du Sud sont remarquablement bien identifiés (dans 96 à 98 % des cas). Les arbres de décision donnent majoritairement la bonne réponse pour les classes représentées par un grand nombre d'instances (standard et Sud). Lorsque l'ensemble restreint d'attributs est utilisé, un grand nombre d'erreurs est dû à l'attribution de la classe standard à des locuteurs d'Alsace, de Belgique et de Suisse. Dans l'autre configuration, la bonne réponse est la plus fréquente pour les Belges et les Suisses, mais une majorité d'Alsaciens ont été identifiés comme Belges, et aucun comme Alsacien.

Classification en trois variétés

Dans un second temps, nous avons cherché à classer nos locuteurs en trois variétés de français. Nous avons vu au chapitre 3 que les auditeurs avaient du mal à différencier les français d'Alsace, de Belgique et de Suisse : il n'est pas absurde de les regrouper en une seule macro-variété, que nous avons appelée Est. De cette manière, les trois classes considérées (Est, Standard et Sud) sont plus équilibrées en termes de nombre d'instances les composant.

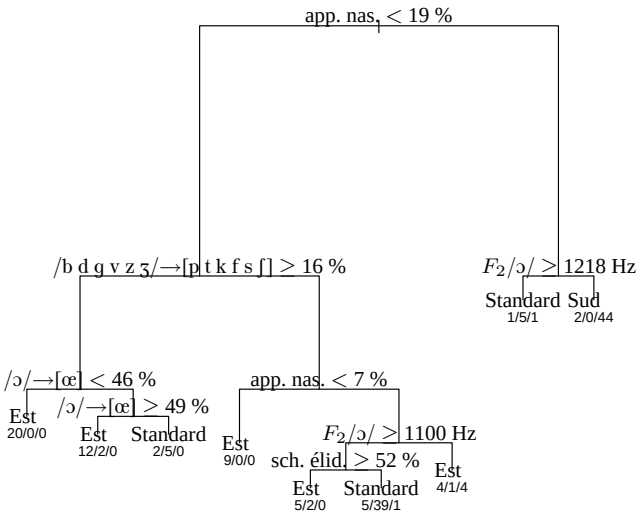
Il n'était pas dit qu'il serait facile de trouver des indices permettant d'identifier à la fois les locuteurs d'Alsace, de Belgique et de Suisse, même si les arbres obtenus dans la section précédente pouvaient le laisser penser (ces trois variétés se retrouvaient dans les mêmes branches de l'arbre). Les taux d'identification présentés dans le tableau 7.3 suggèrent que la tâche est plus simple que la classification en 5 variétés pour les arbres de décision et les SVM. Ces taux, allant jusqu'à 85 % de réponses correctes, sont presque tous supérieurs à ceux que nous avons obtenu pour les 5 variétés.

	ensemble d'attributs	lecture ($\sim$ 3 min)	spontané ($\sim$ 13 min)	tout ($\sim$ 16 min)
Arbres	restreint	68	80	71
	étendu	69	83	69
SVM	restreint	70	79	77
	étendu	73	80	85

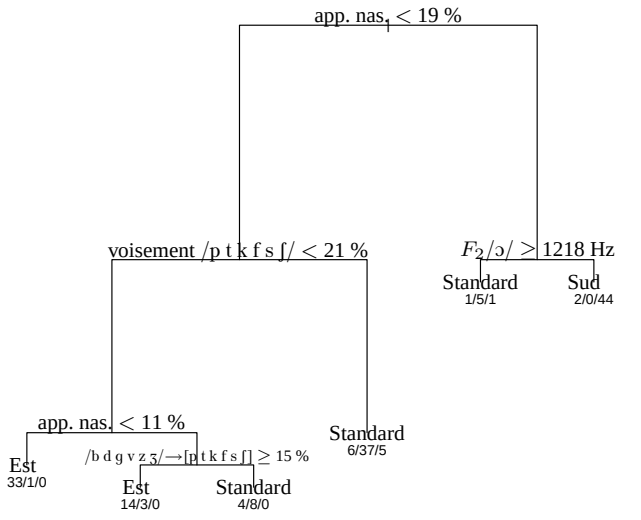
TABLEAU 7.3. Pourcentage d'instances correctement classifiées en 3 variétés pour la lecture, la parole spontanée et ces deux styles confondus, par les arbres de décision et les SVM, avec les ensembles d'attributs restreint et étendu. Les pourcentages sont donnés par rapport à un peu moins de 170 locuteurs. Entre parenthèses, la quantité de parole disponible en moyenne pour chaque locuteur.

Comme précédemment, les taux calculés sur la lecture seule restent les plus bas et l'ensemble étendu donne globalement de meilleurs résultats que l'ensemble restreint. Si globalement les SVM donnent toujours de meilleurs résultats, l'écart avec les résultats donnés par les arbres de décision est moins important que lors de la classification en 5 variétés. Les arbres de décision sont notamment plus performants lorsque seule la parole spontanée est prise en compte.

Les arbres construits avec l'ensemble restreint (figure 7.3(a)) et étendu (figure 7.3(b)) sont plus simples que ceux obtenus pour la classification en 5 variétés. Les attributs distinguant la variété du Sud sont les mêmes que précédemment (ce qui n'est pas incohérent car cette classe est restée inchangée) : les appendices nasaux et le second formant du /ɔ/. Dans l'arbre construit avec l'ensemble restreint d'attributs, des variantes d'alignement sont ensuite mises en jeu pour corroborer le dévoisement des consonnes sonores. L'arbre construit avec l'ensemble étendu isole une partie des locuteurs grâce au taux de voisement des consonnes sourdes : nous avions déjà remarqué les valeurs élevées de ce taux au chapitre 5. Il n'est pas certain qu'il reflète une particularité régionale, mais il permet une bonne séparation entre



(a) ensemble d'attributs restreint



(b) ensemble d'attributs étendu

FIGURE 7.3. Arbres de décision construits à partir de tous les locuteurs du corpus PFC pour une classification en 3 variétés.

l'Est et le standard. Les attributs liés à la prosodie ont complètement disparu dans les deux arbres.

Quel que soit l'ensemble pris en compte, des attributs peu convaincants sont utilisés pour séparer les variétés standard et Est. Dans les deux cas, certaines branches séparent le standard et l'Est avec le taux de voyelles nasales alignées avec un appendice nasal. Dans le cas de l'ensemble restreint d'attributs, l'antériorisation du /ɔ/ est utilisée de façon presque contradictoire (les locuteurs de l'Est sont identifiés par un taux inférieur à 46 % ou supérieur à 49 %). La répartition en 3 variétés est formellement plus équilibrée, mais la classe Est semble peu homogène.

Nous avons également construit les matrices de confusion entre ces 3 variétés : elles sont présentées dans le tableau 7.4 pour les arbres de décision et les SVM. Pour nos deux classifieurs, la réponse donnée est majoritairement la bonne dans tous les cas. La confusion la plus fréquente a lieu entre les locuteurs de l'Est et ceux du français standard, ce qui est cohérent à la fois avec la structure des arbres et avec les observations faites aux chapitres précédents. Les locuteurs du Sud ont été correctement identifiés dans de très nombreux cas (jusqu'à 94 ou 98 % avec les SVM). Il faut également noter la présence de zéros sur la ligne qui leur correspond : ils ont peu été confondus avec les autres variétés. Les arbres de décision et les SVM sont pour cette tâche performants en comparaison avec les auditeurs humains : en effet, pour la tâche de classification en 3 variétés, ces derniers obtiennent de bons taux d'identification pour les variétés Sud et Est (91 et 83 % respectivement), mais identifient le plus souvent les locuteurs de la variété standard comme étant orginaires de l'Est (dans 57 % des cas). Ces résultats ont cependant été influencées par les catégories proposées dans le test perceptif : une pour le standard contre cinq pour l'Est.

	origine \ réponse	Arbres			SVM		
		Standard	Est	Sud	Standard	Est	Sud
restreint	Standard	**67**	24	9	**72**	26	2
	Est	25	**65**	10	23	**68**	8
	Sud	8	8	**84**	0	6	**94**
étendu	Standard	**67**	24	9	**80**	17	4
	Est	35	**62**	3	18	**78**	3
	Sud	20	0	**80**	2	0	**98**

TABLEAU 7.4. Matrice de confusion (%) obtenue par validation croisée *leave-one-out* sur l'ensemble des données PFC classifiées en 3 variétés, en considérant les ensembles restreint et étendu d'attributs, avec les arbres de décision et les SVM.

Que la tâche soit de classifier nos locuteurs en trois ou en cinq variétés, nous avons obtenu des taux d'identification satisfaisants. Des améliorations (paramétrages des classifieurs, sélection automatique des attributs. . .) pourraient bien entendu être apportées, mais étant don-

née la quantité de parole limitée pour certaines classes, nous risquerions de faire du surapprentissage à partir de nos données.

7.2.3 Classification sur le corpus CTS

Dans le corpus CTS, une instance correspond à une intervention d'un locuteur, soit une demi-conversation téléphonique. Nous avons déjà mentionné le fait que, dans ce corpus, un même locuteur peut intervenir dans plusieurs conversations sans que cette information soit codée dans les enregistrements. Avec notre méthode de calcul, un locuteur apparaissant n fois dans des enregistrements distincts correspondra à n instances différentes. De ce fait, nous éviterons d'appliquer la méthode *leave-one-out* sur les données CTS : la classification d'une instance pourrait être biaisée par la présence d'instances correspondant au même locuteur dans l'ensemble d'apprentissage. Il faudrait exclure toutes les instances correspondant à un même locuteur, ce qui reviendrait à mettre en œuvre un mécanisme plus complexe — d'identification du locuteur [Bimbot *et al.*, 2004], par exemple. La technique du *leave-one-out* est donc moins adaptée au corpus CTS qu'au corpus PFC. Il est toutefois intéressant d'appliquer des technique de classification sur ce corpus. Un classifieur entraîné sur les données PFC est-il efficace pour classer les données CTS ? Avant de pouvoir répondre à cette question, nous devons préciser certains points. Le corpus CTS ne contenant parmi les variétés de français étudiées que des données standard, Sud et Alsace, nous pouvons envisager d'entraîner le classifieur avec les données PFC de plusieurs manières. Devons-nous considérer une grande région Est, comprenant l'Alsace, la Belgique et la Suisse ? Faut-il uniquement conserver l'Alsace ?

Nous avons testé ces deux configurations : dans tous les cas, les classifieurs entraînés en considérant une variété Est donnent des taux d'identification inférieurs d'au moins 20 % à ceux que l'on peut entraîner à partir de l'Alsace seule. Cette observation étaye l'hypothèse émise dans la section précédente selon laquelle la variété Est serait peu homogène. Nous avons par conséquent conservé la dernière configuration, qui donne les meilleurs taux d'identification : les résultats sont présentés dans le tableau 7.5.

	ensemble d'attributs	spontané ($\sim$ 13 min)	tout ($\sim$ 16 min)
Arbres	restreint	64	61
	étendu	64	61
SVM	restreint	54	70
	étendu	54	71

TABLEAU 7.5. Pourcentage d'instances du corpus CTS correctement classifiées par des arbres de décision et des SVM entraînés avec les données PFC. Les pourcentages sont donnés par rapport à un peu moins de 600 interventions téléphoniques. Entre parenthèses, la quantité de parole disponible en moyenne pour chaque locuteur ayant servi à entraîner le classifieur.

Étant donné que le corpus CTS ne comporte que des conversations spontanées, devons-nous entraîner notre classifieur uniquement avec de la parole spontanée ? Faut-il au contraire utiliser la totalité de la parole dont nous disposons pour chaque locuteur (y compris la lecture), afin que l'apprentissage exploite davantage de données ? Nous donnons les résultats pour ces deux configurations, car nos deux classifieurs ne se comportent pas de la même manière : les arbres de décision ont de meilleures performances (que les SVM également, de façon intéressante) quand ils sont entraînés sur la parole spontanée uniquement, tandis que les SVM sont plus efficaces sur une plus grande quantité de données.

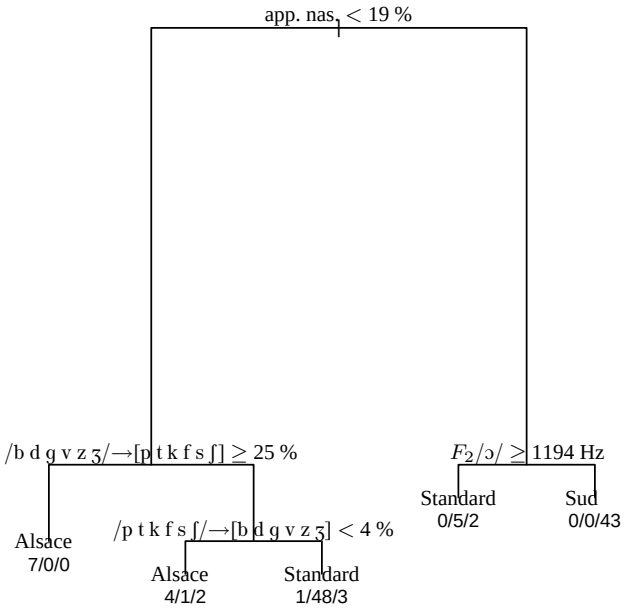


FIGURE 7.4. Arbre de décision construit à partir des locuteurs des variétés standard, Sud et Alsace du corpus PFC pour classer les locuteurs du corpus CTS.

Nous avons tracé l'arbre de décision entraîné avec les deux style de parole dans la figure 7.4. Nous n'avons représenté qu'un seul arbre qui correspond à la fois aux ensembles restreint et étendu d'attributs : ces deux arbres sont en effet identiques. La structure obtenue est assez simple et reste cohérente avec les arbres construits précédemment. Les locuteurs du Sud sont séparés des autres grâce aux appendices nasaux et au second formant du /ɔ/, les Alsaciens par le dévoisement des consonnes sonores, puis par un faible taux de consonnes sourdes alignées comme sonores.

Les matrices de confusion (tableau 7.6) révèlent que de nombreux locuteurs de l'Est et du Sud ont été à tort identifiés comme appartenant à la variété standard, quels que soient l'ensemble d'attributs et le classifieur employés. Dans les cas où cela se produit, la bonne réponse est en deuxième position. Seuls les SVM donnent majoritairement la bonne réponse pour le Sud avec l'ensemble restreint d'attributs. Ils ont été plus efficaces pour cette variété que pour celle d'Alsace, tandis que les arbres de décision ont été plutôt plus performants

	origine \ réponse	Arbres			SVM		
		Standard	Alsace	Sud	Standard	Alsace	Sud
restreint	Standard	**85**	15	0	**91**	5	4
	Alsace	**62**	36	3	**67**	18	15
	Sud	**47**	20	34	41	21	**51**
étendu	Standard	**85**	15	0	**97**	1	2
	Alsace	**62**	36	3	**79**	10	10
	Sud	**47**	20	34	**51**	4	46

TABLEAU 7.6. Matrice de confusion (%) obtenue sur l'ensemble des données CTS classifiées en 3 variétés, en considérant les ensembles restreint et étendu d'attributs, avec les arbres de décision et les SVM.

pour l'Alsace. Les locuteurs du français standard sont, eux, correctement identifiés dans la majorité des cas. Leur classe attire également considérablement l'identification, ce qui peut être causé par l'accent moins marqué des locuteurs du corpus CTS.

Les taux d'identification obtenus ici sont bien évidemment moins élevés que ceux que nous avons mesurés par *leave-one-out* sur le seul corpus PFC. Ils ne dépassent guère les 70 % pour un choix entre 3 classes. Ces différences de scores s'expliquent par le changement de nature des enregistrements (passage de la bande large à la bande téléphonique) et par un moindre contrôle de l'origine des locuteurs dans le corpus CTS... Compte tenu de ces difficultés, les résultats obtenus sont encourageants, nos indices permettant d'identifier l'origine de plus de deux tiers des locuteurs.

7.3 Conclusion

Nous avons utilisé dans ce dernier chapitre deux classifieurs, chacun avec ses avantages (facilité d'interprétation et rapidité pour les arbres de décision, meilleures performances sur nos données, globalement, pour les SVM), pour traiter nos deux corpus. À l'aide de la validation croisée *leave-one-out*, nous avons obtenu des taux d'identification atteignant les 82 et 85 % pour classifier nos locuteurs en respectivement 5 et 3 classes. Ces taux sont remarquablement proches. Le manque d'homogénéité de la classe Est peut expliquer le peu de dégradation observé en passant de 3 à 5 classes. Utilisés pour classifier les données du corpus CTS en 3 classes (standard, Sud et Alsace), les classifieurs entraînés sur le corpus PFC nous ont permis d'obtenir jusqu'à 71 % d'identification.

Diverses améliorations pourraient être apportées pour cette tâche de classification. Avec davantage de données à disposition (même si le volume traité ici n'est pas négligeable), nous pourrions équilibrer le temps de parole par locuteur ainsi que le nombre d'instances par classe. Il faudrait également classifier des données non vues du même type que les données

ayant servi à l'entraînement pour compléter la validation croisée *leave-one-out*. Avec une quantité suffisante de données, nous pourrions mieux entraîner et paramétrer les classifieurs tout en évitant le surapprentissage, et ainsi tenter d'obtenir de meilleurs résultats. Il demeure que les taux d'identification correcte que nous avons obtenus sont prometteurs : ils sont déjà relativement élevés, et plusieurs pistes s'offrent à nous pour essayer de les améliorer.

Nos résultats valident également l'utilité des paramètres que nous avons calculés pour une tâche d'identification de l'accent régional. Nous avons montré que les indices acoustiques mesurés sur nos corpus contribuent relativement bien à identifier automatiquement l'origine géographique des locuteurs.

Les paramètres que nous avons utilisés sont liés à des phénomènes linguistiques : il serait intéressant d'aller plus loin dans le parallèle avec la perception humaine, afin d'étudier les similitudes entre les deux approches linguistique et automatique. Nous pourrions également mettre en relation l'identification des locuteurs avec le degré d'accent qui leur a été attribué lors des tests perceptifs. Les auditeurs se sont-ils appuyés sur les mêmes paramètres que la machine ? Nos indices reflètent-ils vraiment les phénomènes linguistiques dont ils sont censés rendre compte ? Ces dernières questions restent ouvertes et devront faire l'objet d'études plus approfondies.

Notre démarche pourrait être adaptée dans le cadre d'une tâche de transcription enrichie de documents audio assez longs. Elle devrait cependant être adaptée pour traiter des échantillons plus courts.

Conclusion

Tout au long de cette thèse, nous avons travaillé sur deux grands corpus de français comprenant divers accents régionaux : PFC (parole face à face) et CTS (parole téléphonique). À partir de ces enregistrements, nous avons étudié la perception humaine des accents et analysé les caractéristiques de différentes variétés de français — français standard, français du sud de la France, d'Alsace, de Belgique et de Suisse — afin de les modéliser en nous limitant délibérément aux aspects segmentaux et, dans une moindre mesure, suprasegmentaux. Nous nous sommes en grande partie appuyée sur le système d'alignement automatique en phonèmes du LIMSI.

Une partie de nos données a fait l'objet de tests perceptifs (chapitre 3) qui ont permis de montrer quels accents confondent des auditeurs français. Nous avons mis en évidence un continuum perceptif allant du français standard à la Suisse romande et à l'Alsace en passant par la Belgique. Les accents du sud de la France, bien que souvent confondus entre eux, sont bien distingués des autres accents présents dans les tests perceptifs.

Nous avons ensuite caractérisé nos données grâce à des techniques automatiques en employant deux approches complémentaires. Nous avons d'une part mesuré des paramètres acoustiques tels que les formants (chapitre 4), la fréquence fondamentale et l'intensité (chapitre 5) en nous appuyant sur les frontières temporelles des phonèmes données par un alignement standard. Nous avons d'autre part introduit des variantes dans le dictionnaire de prononciation utilisé lors de l'alignement (chapitre 6), afin d'observer les variantes choisies par le système. Ces deux méthodes nous ont permis de mettre en évidence un certain nombre de traits de prononciation, sur lesquels nous reviendrons.

Finalement, nous avons appliqué des techniques de classification automatique à nos données : arbres de décision et SVM (chapitre 7). Les indices motivés linguistiquement que nous avons utilisés ont permis d'obtenir des taux d'identification satisfaisants : jusqu'à plus de 80 % sur 3 ou 5 variétés du corpus PFC en validation croisée *leave-one-out* et jusqu'à plus de 70 % sur 3 variétés du corpus CTS avec un classifieur entraîné sur les données PFC.

Notre travail a dégagé un certain nombre de traits caractéristiques de variétés régionales de français dans nos corpus, qui sont récapitulés dans le tableau 7.7. Nous avons montré que ces traits peuvent être mesurés de manière automatique à partir du signal de parole. Nous avons travaillé sur les aspects segmentaux (les voyelles et certaines consonnes) et supra segmentaux (la prosodie). Les locuteurs du Sud présentent plusieurs traits, relativement robustes, qui leur sont propres : ils réalisent davantage de schwas et de voyelles nasales partiellement dénasalisées et suivies d'un appendice consonantique nasal. De plus, le /ɔ/ tend,

chez eux, à être prononcé d'une manière proche du [o]. Tous ces traits sont cités comme caractéristiques du Sud dans différentes études linguistiques (chapitre 1). Les locuteurs du français standard ont pour leur part tendance à prononcer le /ɔ/ de manière antériorisée, se rapprochant du [œ]. Ce phénomène apparaît à travers les mesures de formants aussi bien qu'avec les variantes de prononciation autorisant l'alignement /ɔ/→[œ]. Nous avons observé en Alsace un dévoisement de consonnes sonores /b d g v z ʒ/, ainsi qu'une tendance à l'accentuation initiale. Nos Belges ont tendance à allonger la penultième syllabe avant une pause et à prononcer plus fréquemment que les autres locuteurs des /ʁ/ proches de la jota [x]. Nos locuteurs suisses présentent une accentuation initiale ; ils allongent également les deux dernières syllabes avant une pause.

	trait	méthode	
Standard	antériorisation du /ɔ/	F_2 /ɔ/ % /ɔ/→[œ]	*(formants)* *(variantes)*
Sud	voyelles nasales avec un appendice consonantique nasal	% VN → V+CN	*(variantes)*
	schwas élidés	% schwas élidés	*(variantes)*
	rapprochement /ɔ/ /o/	F_2 /ɔ/ % /ɔ/ → [o]	*(formants)* *(variantes)*
Alsace	dévoisement des consonnes sonores	% /b d g v z ʒ/→[p t k f s ʃ] % voisement /b d g v z ʒ/	*(variantes)* *(voisement)*
	allongement initial	Δintensité, Δdurée	*(prosodie)*
	rapprochement /ɔ/ /o/ ?	F_2 /ɔ/ : ? % /ɔ/→[o]	*(formants)* *(variantes)*
Belgique	réalisation /ʁ/	/ʁ/→[x] % voisement /ʁ/	*(variantes)* *(voisement)*
	allongement pénultième	$\Delta\text{durée}_f$	*(prosodie)*
Suisse	accentuation initiale	ΔF_0, Δintensité, durée attaque	*(prosodie)*
	allongement pénultième et finale	$\Delta\text{durée}_f$, $\Delta\text{durée}_p$	*(prosodie)*

TABLEAU 7.7. Récapitulatif des traits de prononciation caractéristiques de chaque variété de français et des méthodes ayant permis de les mettre en évidence.

En règle générale, les différences que nous avons pu observer au niveau prosodique sont moins marquées que les différences au niveau segmental. Elles apparaissent également moins souvent dans les arbres de décision appris automatiquement. Au contraire, les traits que nous avons mesurés pour le Sud semblent très robustes et permettent une très bonne distinction de cette variété de français. Nous n'avons pas mis en relation le caractère plus ou moins marqué de certains traits ainsi que les taux d'identification automatique avec le degré

d'accent attribué aux échantillons de parole : le rapport entre les deux pourrait toutefois être instructif.

Nous aurions pu poursuivre l'étude de certains phénomènes avec des techniques déjà employées dans notre travail. D'après certains auteurs, le français du sud de la France présente une prosodie particulière. Bien que la réalisation de plus nombreux schwas soit déjà l'indice d'une prosodie différente pour cette variété, les patrons prosodiques que nous avons étudiés ne permettent pas par exemple d'observer une différence au niveau de F_0 entre français standard et méridional. Comme nous disposions déjà de beaucoup d'indices pour cette dernière variété, nous n'avons pas approfondi cette question, mais il pourrait être pertinent de tenter de trouver un patron prosodique caractéristique du sud de la France. Par ailleurs, l'étude de variantes de prononciation comme /e/∼/ɛ/ nous semble intéressante notamment pour les locuteurs suisses.

Plusieurs possibilités s'offrent pour poursuivre l'étude des accents régionaux : considérer des points d'enquête supplémentaires, étudier plus finement ceux que nous avons étudiés, inclure d'autres variétés de français parlées hors d'Europe, en Amérique, en Afrique et dans l'Océan Indien.

L'ajout de points d'enquête se heurte toutefois à un relatif manque de données disponibles pour certaines variétés. Nous sommes bien consciente d'avoir commis quelques abus de langage en parlant par exemple des Suisses, quand seulement une douzaine de locuteurs de Nyon ou du canton de Vaud sont représentés. Mais ceci est inhérent à nombre d'études sur la variation. La nôtre manipule globalement une grande quantité de données que seul le traitement automatique de la parole a permis d'aborder.

En analysant d'autres traits de prononciation, nous pourrions tenter de différencier certains points d'enquêtes au sein des variétés que nous avons définies (Sud, standard ou Belgique). Une analyse plus fine peut toutefois trouver des limites, dans le cas où la différence entre deux accents s'effectuerait sur la prononciation d'un mot peu fréquent, qui n'apparaîtrait pas dans un corpus fermé. Nous pourrions également faire émerger de nouvelles différence en mettant en place d'autres tests perceptifs. Il pourrait également être intéressant de mener des tests perceptifs en informant les auditeurs des caractéristiques que nous avons mis en évidence pour chaque variété et de vérifier s'ils se montreraient alors plus performants.

Nous nous sommes focalisée sur l'étude de la variation régionale, mais d'autres sources de variation apparaissent également dans nos enregistrements. La divergence indéniable entre langue écrite et langue parlée (variation diamésique), comme entre parole lue et parole spontanée (variation diaphasique), a fait l'objet d'un grand nombre d'études [Léon, 1993; Adda-Decker et Lamel, 1999; Boula de Mareüil et Adda-Decker, 2002; Boula de Mareüil *et al.*, 2005]. Dans nos travaux, nous avons examiné différents styles de parole (lecture et parole spontanée) représentés dans le corpus PFC, entre lesquels nous avons pu observer des différences : lors de l'étude des variantes de prononciation, nous avons observé davantage de réalisations de variantes canoniques dans la lecture. Il est difficile d'intégrer la dimension sociale en plus de la variation régionale sur nos corpus : elle n'est pas centrale dans le corpus PFC et elle est complètement absente du corpus CTS. Nous n'avons pas mesuré l'importance relative des frontières socioculturelles par rapports aux frontières géographiques (ces dernières étant traditionnellement considérées comme plus importantes en français [Walter, 1988]). Pourtant, les méthodes que nous avons appliquées aux accents régionaux pourraient

être étendues aux accents sociaux.

Les échantillons de parole que nous avons soumis à des tests perceptifs étaient d'une dizaine de secondes ; ceux qui ont fait l'objet d'expériences en identification automatique étaient plus longs, de quelques minutes. Il serait opportun d'appliquer les protocoles en vigueur dans les grandes campagnes d'évaluation en identification des langues (soit environ 45 secondes [Pellegrino et André-Obrecht, 2000]). Les données manquent pour ce faire : les grandes campagnes d'évaluation ne s'ouvrent que depuis peu à l'identification de variétés de langues, dans la mesure où l'identification des langues fonctionne bien pour des échantillons de parole assez longs (par exemple 1 min). Cette tâche est bien plus difficile à mettre en œuvre, et requiert — dans le cas où les variétés considérées sont situées à l'intérieur d'un même pays — une connaissance plus fine des locuteurs. Dans une telle perspective, notre démarche devrait très certainement être adaptée, mais la comparaison entre les performances de l'humain et de la machine serait ainsi plus juste. Une autre problématique en lien avec nos travaux est celle de la transcription enrichie (comme dans [Galliano *et al.*, 2006]) : en plus des mots transcrits, des informations sur l'identité (par exemple le sexe) ou l'origine (l'accent) du locuteur doivent être déterminées.

Les indices acoustiques permettant de distinguer automatiquement différentes variétés régionales de français peuvent trouver des applications directes. Le taux d'erreur des systèmes de reconnaissance de la parole augmente souvent lorsqu'ils doivent traiter de la parole avec accent : l'utilisation de nos indices pourrait contribuer diminuer ces taux. L'enrichissement des dictionnaires de prononciation est une première piste à explorer dans cette voie.

Annexes

Annexe A

Texte PFC

Le Premier Ministre ira-t-il à Beaulieu ?

Le village de Beaulieu est en grand émoi. Le Premier Ministre a en effet décidé de faire étape dans cette commune au cours de sa tournée de la région en fin d'année. Jusqu'ici les seuls titres de gloire de Beaulieu étaient son vin blanc sec, ses chemises en soie, un champion local de course à pied (Louis Garret), quatrième aux jeux olympiques de Berlin en 1936, et plus récemment, son usine de pâtes italiennes. Qu'est-ce qui a donc valu à Beaulieu ce grand honneur ? Le hasard, tout bêtement, car le Premier Ministre, lassé des circuits habituels qui tournaient toujours autour des mêmes villes, veut découvrir ce qu'il appelle « la campagne profonde ».

Le maire de Beaulieu – Marc Blanc – est en revanche très inquiet. La cote du Premier Ministre ne cesse de baisser depuis les élections. Comment, en plus, éviter les manifestations qui ont eu tendance à se multiplier lors des visites officielles ? La côte escarpée du Mont Saint-Pierre qui mène au village connaît des barrages chaque fois que les opposants de tous les bords manifestent leur colère. D'un autre côté, à chaque voyage du Premier Ministre, le gouvernement prend contact avec la préfecture la plus proche et s'assure que tout est fait pour le protéger. Or, un gros détachement de police, comme on en a vu à Jonquière, et des vérifications d'identité risquent de provoquer une explosion. Un jeune membre de l'opposition aurait déclaré : « Dans le coin, on est jaloux de notre liberté. S'il faut montrer patte blanche pour circuler, nous ne répondons pas de la réaction des gens du pays. Nous avons le soutien du village entier. » De plus, quelques articles parus dans La Dépêche du Centre, L'Express, Ouest Liberté et Le Nouvel Observateur indiqueraient que des activistes des communes voisines préparent une journée chaude au Premier Ministre. Quelques fanatiques auraient même entamé un jeûne prolongé dans l'église de Saint Martinville.

Le sympathique maire de Beaulieu ne sait plus à quel saint se vouer. Il a le sentiment de se trouver dans une impasse stupide. Il s'est, en désespoir de cause, décidé à écrire au Premier Ministre pour vérifier si son village était vraiment une étape nécessaire dans la tournée prévue. Beaulieu préfère être inconnue et tranquille plutôt que de se trouver au centre d'une bataille politique dont, par la télévision, seraient témoins des millions d'électeurs.

Annexe B

Énoncés spontanés retenus pour les tests perceptifs

B.1 Premières expériences

Le nom du fichier .wav indique l'identité du locuteur : les deux premières lettres codent la région (no = Vendée, nn = Normandie, ne = Suisse romande, so = Pays basque, ss = Languedoc, se = Provence), la troisième la tranche d'âge (j = jeune, m = moyen, v = vieux), la quatrième le style de parole (ici toujours s pour parole spontanée) et la cinquième le sexe (h = homme, f = femme).

ssvsh.wav c'est essayer de faire une une liste d'union, euh nous on ne veut pas se présenter à tout prix contre vous, on voudrait se présenter tous les trois avec vous, voilà prenez nous tous les trois

ssvsf.wav parce que c'était une portion hein c'était pas grand chose c'était cent ou cent-cinquante grammes de viande, ou ou le pain pour le pain aussi nous avions des tickets

ssmsh.wav je je je fais une exposition de vieux outils mais pas pas pour en tirer profit pour euh pour faire visiter c'est personnel c'est... c'est c'est c'est pour mon plaisir personnel

ssmsf.wav elle elle avait une soif de s'instruire constamment, dès qu'elle allumait la télévision et qu'elle entendait des mots qu'elle ne connaissait pas trop elle me demandait vite

ssjsh.wav faut faire des analyses de sol aussi, c'est eux qui le font, euh s'il y a trop de de métaux lourds dans le sol trop de plomb on n'a pas droit au bio

ssjsf.wav là je me suis rendu compte que j'étais vraiment passée dans un autre système au lycée eh bé on nous assistait un petit peu entre guillemets je dirais et là c'est le contraire il faut réagir différemment

sovsh.wav parce que eux ils pensaient d'ailleurs que c'était un dialecte alors que c'est pas un dialecte il y a une culture très très ancienne il y a une il y a une écriture il y a une littérature qui date de euh de Ronsard à la même époque

sovsf.wav et une soeur qui s'appelait Eugénie, qui elle s'occupait de de la maison des travaux de du ménage quoi

somsh.wav ce qu'on avait appris un peu avant on faisait du français on faisait de des langues étrangères on faisait de la comptabilité on faisait enfin pour acquérir ce métier je pense que le mieux c'est c'est le terrain hein je pense

somsf.wav il y a toute la partie qui est pas qui est pas judiciaire quand les gens euh nous sollicitent pour des changements de de régimes matrimoniaux, pour des adoptions d'enfants et caetera

sojsh.wav ce coup de pied là eh bé ce sera pas une bonne génération ils sont déjà descendus euh dans la rue une première fois ils seront pet-être capables de se mobiliser euh d'autres pour d'autres occasions

sojsf.wav et donc euh bon la maternelle je pense que c'est comme enfin j'ai pas trop de souvenirs non plus mais je sais que on était assez proches des euh des institutrices

sevsh.wav et donc il y avait un médecin et un prêtre, deux frères, et les deux frères descendaient avec une domesticité quelconque à l'époque euh les prêtres avaient euh leur bonne

sevsf.wav cette année ils ont dit on veut une volaille parce que l'année dernière j'avais fait des des de la queue de langouste alors ils m'ont dit non non on veut une volaille cette année

semsh.wav tu peux pas dire oui il l'est ou non il l'a pas, alors tu dis oui il est peut-être que oui là il est peut-être que non ben alors après on peut jouer sur les mots les gens vont le savoir ils te ils viennent te voir ils te rapellent est-ce que

semsf.wav ce qui était un peu un peu pénible mais enfin bon en rentrant chez moi je mettais les pieds sous la table donc pas de problème j'avais le temps de rentrer tranquillement à la maison

sejsh.wav disons qu'on a un programme assez chargé, l'avantage c'est que que qu'on va jamais en cours parce que de toutes façons les profs de la fac c'est ce qu'ils nous disent de toutes façons le cours n'est pas complet

sejsf.wav euh à la fin de la classe verte il y avait quelqu'un qui lui avait demandé qu'est-ce que tu préfères qu'est-ce que tu as préféré faire pendant cette semaine elle avait dit dormir

novsh.wav on couchait n'importe où à la découverte sur de pierres sur euh ou dans dans dans un local désaffecté ou n'importe

novsf.wav on avait on avait de la famille oui elle avait une tante bon ça a sûrement aidé à ce que à ce que ça se fasse comme ça mais c'était euh c'était pas évident non plus hein

nomsh.wav quand mon père a pris sa retraite eh ben j'ai repris la suite euh de la forge, alors j'ai été forgeron à mon compte pendant euh pff dix-sept ans

nomsf.wav et avec le père et la mère et après c'est le fils qui a pris la suite, et ce fils qui était un boulanger qui était boulanger il a repris la suite de son père avec sept enfants à la maison

nojsh.wav et après je suis arrivé dans la classe euh j'étais un des plus vieux, quand j'étais en diesel j'étais un des plus vieux parce que j'avais euh déjà d'une j'avais redoublé et puis en plus j'avais fait deux BEP

nojsf.wav et ce que je voudrais faire c'est euh faire tout ce qui est classe verte classe nature, emmener les enfants euh si tu veux en classe découverte montrer tout ce qui est faune flore tout ça

nnvsh.wav je suis resté en très bonne relation avec son fils, je vais le voir demain midi, et il m'a dit il il me dit toujours Roger vous êtes mon père spirituel

nnvsf.wav et puis du de de la directrice de les euh de l'école qui était là actuellemnt qui était peut-être un peu momolle mais qui avait quand même euh qui savait remettre les choses en place

nnmsh.wav avec lequel j'ai été mon professeur d'anglais avec lequel j'avais beaucoup de lacunes, et nous avions quelques différents tous les deux et je le regrette bien aujourd'hui

nnmsf.wav parce que j'avais un professeur de technologie euh qui nous permettait de quitter le cours dès qu'on avait terminé donc euh mon frère étant né un matin juste avant que je parte à l'école

nnjsh.wav et là bah cette année au premier semestre en en première année de fac j'avais pris portugais et ça me plaisait pas donc bah j'ai arrêté j'ai j'ai préféré garder l'anglais parce que

nnjsf.wav beaucoup dans l'éco aussi enfin tout ce qui est gestion tout ça, mais moi je veux faire un je veux aller jusqu'au master en fait soit réintégrer la filière droit

nevsh.wav que au point de vue oral, l'expression n'est plus quelque chose qui est important chez les jeunes, la preuve elle en est la téléphonie mobile

nevsf.wav bon moi je suis toujours restée euh très active en comptabilité ou je peux faire des ? ? ? ici à la maison j'aime beaucoup l'extérieur le jardin les fleurs

nemsh.wav puis ils installaient toujours la ménagerie, on était quasiment dans la ménagerie avec notre école, puis là il y avait le l'enclos du rhinocéros

nemsf.wav autrement mon école moi j'ai toujours eu euh un bon caractère donc j'ai toujours euh j'ai toujours bien suivi et puis euh j'ai encore gardé encore à l'heure actuelle j'ai encore des amis

nejsh.wav ou bien des fois quand il y a des coups de vent tout d'un coup ça souffle très fort, et puis il y a des gens avec les voiliers comme ça ils arrivent pas à rentrer dans les ports

nejsf.wav parce qu'on est quand même bien ici malgré tout hein, et puis de voir ces gens qui meurent de faim qui ont pas de quoi donner à manger à leurs enfants ça me travaillerait trop

B.2 Deuxième expérience

Nous donnerons ici uniquement la transcription des stimuli non utilisés lors de la première série d'expériences. Les noms des fichiers .mp3 indiquent l'identité du locuteur de la même

manière que précédemment (aa = Boersch (Alsace), bg = Gembloux (Belgique), bl = Liège (Belgique), bt = Tournai (Belgique)).

aavsh.mp3 euh jusqu'à l'âge de trente ans on a fait des et encore actuellement on joue des fois y'a deux orchestres maintenant y'a pas seulement la musique populaire comme on dit

aavsf.mp3 moi j'étais en bas et puis il y avait euh mon papa ou ou qui on amenait une bassine pleine d'eau

aamsh.mp3 il n'était pas question de rentrer à seize heures de l'école et de poser le cartable d'abord on faisait les devoirs après on goutait et après on allait gambader

aamsf.mp3 et elles avaient de grands parents plus âgés vous savez euh par exemple la maman de de ma copine a quatre-vingts ans

bgvsh.mp3 ah il y a les beaux côtés du métier et il y a les laids côtés mais dans l'ensemble je n'ai pas à me plaindre de mon métier parce que en somme

bgvsf.mp3 c'est quand même déjà assez grand hein ici aussi hein et alors je voudrais bien repeindre en coquille d'œuf comme ça parce que j'étais j'ai acheté des tentures en toile de Jouy

bgmsh.mp3 donc là malgré tout c'est ça c'est beaucoup plus dur ça encore on ne fait que grimper grimper grimper donc pendant quatre cinq heures de suite et puis il faut refaire l'inverse il faut redescendre

bgmsf.mp3 et puis alors donc je m'occupais de du secrétariat du de la préparation des retraites donc on part une fois par an du vendredi au dimanche soir

blvsh.mp3 la filière où il est finalement devenu employé où il prenait note des commandes il faisait les factures etc. pour le grossiste et c'est ça qui leur a donné l'idée d'ouvrir un commerce de quincaillerie puisque mon père avait toujours

blvsf.mp3 je sais que quand il y a eu des inondations mais ne me demande pas en quelle année qu'ils ils habitaient au bord de l'eau avant qu'on ne fasse les digues toutes les années ils étaient inondés et c'est de ça qu'ils ont déménagé

blmsh.mp3 et puis j'ai fait une année sabbatique et puis j'ai fait mes études d'instituteur en deux ans

blmsf.mp3 il lisait beaucoup des des revues là-dessus il est allé dans les caves puis a fait des concours puis il est allé suivre euh des cours d'IJnologie

btvsh.mp3 euh on a des jeunes qui de plus en plus de jeunes qui écrivent vraiment très bien qui dont on est très contents qui reviennent donc c'est que ça leur plait

btvsf.mp3 les mitrailleurs ils étaient à cheval entre le troisième chasseur et les artilleurs tu vois donc ils étaient vraiment à part

btmsh.mp3 quand je suis rentré du service militaire je me suis réinscrit comme demandeur d'emploi en espérant trouver dans l'enseignement parce que ça me plaisait j'aimais j'aimais bien ça d'ailleurs j'aimais autant j'aimais bien tout

btmsf.mp3 après les les plus grands bébés il a fallu aussi continuer à travailler avec eux donc on a enchaîné avec les cours de natation pour les pour les bébés et puis très vite aussi on a eu de la demande pour les le travail en prénatal

Annexe C

Locuteurs utilisés lors des tests perceptifs

origine	sexe	âge	niveau d'études (âge de fin d'études si indiqué)	profession
Treize-Vents	M	22	BEP automobile, BEP carrossier	Monteur d'options sur bateaux
		59	Scolarisé jusqu'au tertiaire	Forgeron
		87	Études secondaires incomplètes (moins de 14 ans)	Retraité - coiffeur
	F	20	BEP vente (16 ans)	Serveuse
		56	Non renseigné	Aide-soignante
		62	École de coiffure à Paris	Retraitée - coiffeuse
Brécey	M	19	Bac + 2 en cours (1re année DEUG espagnol)	Étudiant
		42	BEPC, CAP ébéniste (moins de 14 ans)	Antiquaire
		81	Certificat d'études	Retraité - courtier en bestiaux
	F	20	Bac + 2 en cours (1re année DEUG AES)	Étudiante
		42	Bac G2	Secrétaire commerciale
		69	Certificat d'études (moins de 14 ans)	Retraitée - agricultrice

origine	sexe	âge	niveau d'études (âge de fin d'études si indiqué)	profession
Nyon	M	28	Apprentissage d'ébéniste (de 15 à 19 ans)	Ébéniste
		44	Apprentissage d'employé de commerce (19 ans)	Employé de commerce dans une banque
		69	Apprentissage de comptable, bac, diplôme d'ingénieur agronome (64 ans), doctorat (68 ans)	Retraité - docteur en science
	F	29	Apprentissage d'employé de commerce (18 ans)	Mère au foyer employée de commerce
		52	Apprentissage d'employé de commerce (19 ans)	Secrétaire en formation
		64	École secondaire	Mère au foyer - employée de bureau
Biarritz	M	30	Formation ingénieur (23 ans), préparation CAPES	Enseignant
		37	CAP, BEP, BTHD	Garçon de café
		60	Médecine + spécialité (27 ans), DEUG littéraire	Retraité médecin anesthésiste
	F	25	IUT Mesures Physique, IUT informatique (23 ans)	Analyste programmeur
		37	DESS de droit des affaires et fiscalité (23 ans)	Avocat indépendant
		91	Brevet supérieur (18 ans)	Retraitée - institutrice
Douzens	M	21	Première année des beaux arts	Étudiant
		42	N'a pas eu le certificat d'études (16 ans)	Cantonnier à la mairie du village
		76	Brevet d'études	Retraité - courtier en vin, viticulteur
	F	23	Études de lettres modernes (licence)	Étudiante, contractuelle au rectorat
		48	Bac	Femme de ménage chez des particuliers
		75	Brevet d'études	Mère au foyer

origine	sexe	âge	niveau d'études (âge de fin d'études si indiqué)	profession
Marseille	M	22	Étudiant en médecine	Étudiant
		53	Maîtrise d'histoire et géographie (23 ans)	Cadre de direction à la caisse d'épargne
		72	Brevet des collèges puis équivalence du bac pour faire une école de génie civil	Ingénieur d'études puis directeur technique dans le BTP
	F	17	Étudiante en sciences	Étudiante
		44	Maîtrise de droit, diplôme des assurances (23 ans)	Chargée de clientèle
		73	BE et secrétariat (18 ans)	Fonctionnaire du cadastre
Boersch	M	56	*Non renseigné*	Employé administratif
		75	*Non renseigné*	Employé administratif
	F	35	*Non renseigné*	Employé et assimilé
		76	*Non renseigné*	Femme au foyer
Gembloux	M	45	Enseignement secondaire technique (électromécanique)	Militaire
		76	*Non renseigné*	*Non renseigné*
	F	50	Régendat en langues germaniques	Employée de banque
		63	Lycée	Commerçante
Liège	M	40 61	*Non renseigné*	
	F	35 75	*Non renseigné*	
Tournai	M	43 77	*Non renseigné*	
	F	46 82	*Non renseigné*	

Annexe D

Formants

D.1 Valeurs utilisées pour filtrer les formants

D'après [Gendrot et Adda-Decker, 2005].

	F_1		F_2				F_3	
	femmes	hommes	femmes		hommes		femmes	hommes
	$<$	$<$	$>$	$<$	$>$	$<$	$>$	$>$
i	900	750	1600	3100	1500	2500	2500	2000
y	900	800	1400	2800	1300	2200	1800	1700
e	900	800	1400	3000	1100	2400	2200	2000
ɛ	1100	1000	1400	2700	1200	2300	2000	2000
a	1100	1000	900	2300	800	2300	1900	1800
œ	1100	1000	800	2400	800	2000	2000	2000
ø	1000	900	700	2300	700	2000	1800	1700
ɔ	1000	900	600	2000	600	1800	2000	1500
o	1000	900	600	1600	600	1600	2100	1500
u	1000	900	400	1500	400	1500	1800	1400

D.2 Valeurs des formants

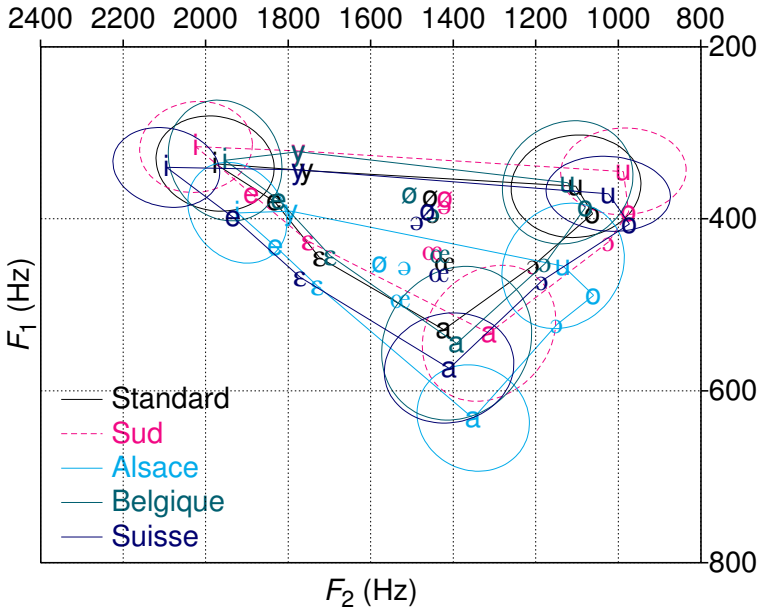
		hommes					femmes				
		Std	Sud	Als	Bel	Sui	Std	Sud	Als	Bel	Sui
a	F_1	520	526	628	540	568	585	637	749	613	636
	F_2	1433	1328	1359	1391	1426	1642	1511	1658	1597	1587
	F_3	2405	2373	2468	2433	2416	2669	2565	2681	2685	2637
ɛ	F_1	436	421	477	438	459	484	490	549	485	504
	F_2	1711	1737	1731	1684	1766	1993	2031	2049	1965	1974
	F_3	2416	2371	2527	2443	2468	2669	2599	2760	2724	2749
e	F_1	377	374	430	380	402	420	427	485	423	435
	F_2	1811	1865	1829	1807	1916	2147	2192	2176	2131	2177
	F_3	2461	2411	2564	2477	2517	2698	2641	2808	2766	2767
i	F_1	329	321	393	331	346	375	378	471	382	379
	F_2	1982	2018	1924	1948	2081	2232	2297	2216	2246	2329
	F_3	2532	2494	2632	2546	2564	2783	2748	2798	2807	2802
ɔ	F_1	437	417	515	443	464	477	475	583	483	500
	F_2	1196	1024	1153	1179	1185	1338	1107	1331	1303	1309
	F_3	2391	2402	2453	2430	2440	2667	2664	2685	2689	2699
o	F_1	385	388	484	384	408	429	435	553	437	442
	F_2	1051	974	1067	1090	993	1147	1035	1223	1145	1053
	F_3	2416	2432	2451	2431	2490	2696	2687	2686	2724	2708
u	F_1	349	344	448	357	371	389	379	520	396	402
	F_2	1083	1003	1138	1136	1035	1154	1078	1249	1142	1083
	F_3	2427	2409	2472	2445	2468	2680	2635	2664	2712	2691
ə	F_1	389	383	453	389	406	423	429	493	424	443
	F_2	1440	1411	1519	1445	1502	1649	1585	1755	1634	1651
	F_3	2417	2385	2478	2427	2439	2677	2626	2693	2706	2711
ø	F_1	370	367	450	366	392	414	423	477	411	417
	F_2	1456	1412	1580	1511	1497	1672	1615	1823	1745	1652
	F_3	2362	2344	2453	2400	2373	2628	2613	2643	2621	2633
y	F_1	331	320	389	322	344	372	367	454	370	367
	F_2	1765	1773	1792	1775	1783	2009	2023	1992	2039	1991
	F_3	2403	2366	2483	2429	2402	2673	2614	2650	2663	2669

TABLEAU D.1. Valeurs moyennes (en Hz) des 3 premiers formants des 10 voyelles orales pour les hommes et les femmes, pour 5 variétés de français du corpus PFC. Les formants sont extraits avec PRAAT, avec trois valeurs par voyelle filtrées avant le calcul de la moyenne.

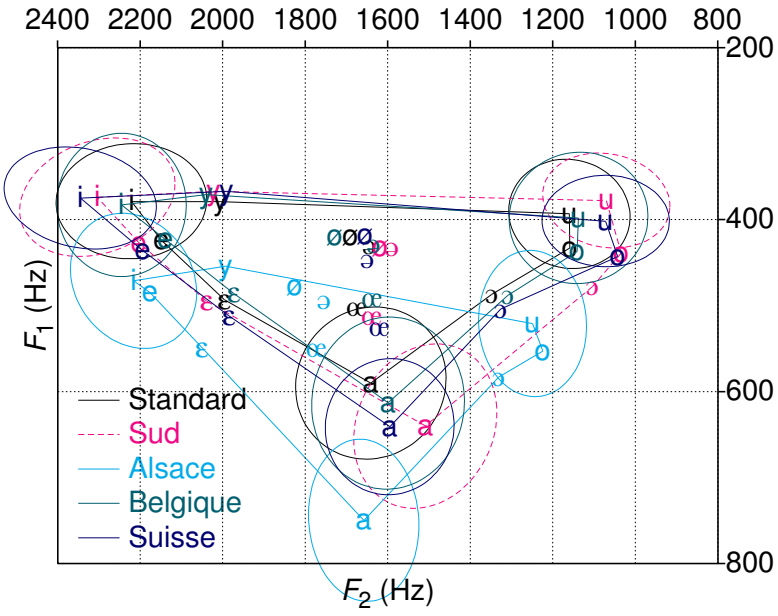
		hommes			femmes		
		Standard	Sud	Alsace	Standard	Sud	Alsace
a	F_1	571	581	593	669	676	703
	F_2	1382	1356	1333	1613	1604	1544
	F_3	2400	2428	2341	2661	2649	2633
ɛ	F_1	512	507	504	569	558	554
	F_2	1635	1656	1603	1946	1959	1950
	F_3	2410	2442	2356	2683	2683	2689
e	F_1	434	431	432	477	476	481
	F_2	1771	1791	1730	2115	2135	2138
	F_3	2458	2488	2410	2732	2749	2749
i	F_1	415	414	409	458	464	458
	F_2	1879	1903	1843	2182	2181	2203
	F_3	2522	2545	2503	2753	2763	2761
ɔ	F_1	519	517	510	583	572	571
	F_2	1203	1108	1138	1347	1274	1238
	F_3	2380	2430	2325	2673	2688	2659
o	F_1	468	476	466	511	517	516
	F_2	1081	1053	1078	1177	1172	1163
	F_3	2410	2458	2342	2703	2715	2692
u	F_1	465	466	453	480	497	484
	F_2	1108	1083	1081	1171	1156	1156
	F_3	2397	2435	2360	2664	2677	2685
ə	F_1	483	470	492	535	519	537
	F_2	1483	1453	1479	1711	1671	1701
	F_3	2407	2431	2374	2684	2701	2692
ø	F_1	448	453	435	489	495	486
	F_2	1422	1407	1415	1642	1627	1664
	F_3	2364	2396	2292	2626	2663	2611
y	F_1	455	456	429	494	504	471
	F_2	1739	1762	1714	1975	1999	1975
	F_3	2410	2421	2380	2671	2690	2659

TABLEAU D.2. Valeurs moyennes (en Hz) des 3 premiers formants des 10 voyelles orales pour les hommes et les femmes, pour 3 variétés de français du corpus CTS. Les formants sont extraits avec PRAAT, avec trois valeurs par voyelle filtrées avant le calcul de la moyenne.

D.3 Triangles vocaliques



(a) Triangles vocaliques des hommes



(b) Triangles vocaliques des femmes

FIGURE D.1. Triangles vocaliques des hommes et des femmes du corpus PFC pour les locuteurs du français standard, du sud, d'Alsace, de Belgique et de Suisse. Les ellipses sont réglées à 20 % des occurences.

Liste des tableaux

Table des figures

Publications de l'auteur

[1] C. Woehrling et P. Boula de Mareüil. Identification d'accents régionaux en français : perception et analyse. *Revue PArole*, 37 :25–65, 2006.

[2] C. Woehrling et P. Boula de Mareüil. Identification of regional accents in French : perception and categorization. In *9th International Conference on Spoken Language Processing*, pages 1511–1514, Pittsburgh, 2006.

[3] P. Boula de Mareüil, M. Adda-Decker et C. Woehrling. Analysis of oral and nasal vowel realisation in northern and southern French varieties. In *16th International Congress of Phonetic Sciences*, pages 2221–2224, Saarbrücken, 2007.

[4] C. Woehrling et P. Boula de Mareüil. Comparing Praat and Snack formant measurements on two large corpora of northern and southern French. In *8th Annual Meeting of the International Speech Communication Association*, pages 1006–1009, Anvers, 2007.

[5] C. Woehrling et P. Boula de Mareüil. Comparaison entre l'extraction de formants par Praat et Snack sur deux grands corpus de français du nord et du sud. In *7es Journées Jeunes Chercheurs en Parole*, pages 152–155, Paris, 2007.

[6] C. Woehrling, P. Boula de Mareüil et M. Adda-Decker. Aspects prosodiques du français parlé en Alsace, Belgique et Suisse. In *27es Journées d'Étude sur la Parole*, pages 1586–1589, Avignon, 2008.

[7] C. Woehrling, P. Boula de Mareüil, M. Adda-Decker et L. Lamel. A corpus-based prosodic study of Alsatian, Belgian and Swiss French. In *9th Annual Meeting of the International Speech Communication Association*, pages 780–783, Brisbane, 2008.

[8] P. Boula de Mareüil, B. Vieru-Dimulescu, C. Woehrling et M. Adda-Decker. Accents étrangers et régionaux en français. Caractérisation et identification. *Traitement Automatique des Langues*, 49(3), 2009.

[9] P. Boula de Mareüil, C. Woehrling et M. Adda-Decker. Apports du traitement automatique à une approche linguistique de la variation régionale dans la parole. In *Méthodes et outils pour l'analyse phonétique des grands corpus oraux*, éditeur N. Nguyen. Hermès, Paris. à paraître.

Bibliographie

[Adank, 2003] P.M. Adank. *Vowel normalization : a perceptual acoustic study of Dutch vowels*. Thèse de Doctorat, Radboud University Nijmegen, 2003.

[Adda-Decker *et al.*, 1999] M. Adda-Decker, P. Boula de Mareüil, et L. Lamel. Pronunciation variants in French : schwa & liaisons. In *14th International Congress of Phonetic Sciences*, pages 2239–2242, San Francisco, 1999.

[Adda-Decker *et al.*, 2005] M. Adda-Decker, P. Boula de Mareüil, G. Adda, et L. Lamel. Investigating syllabic structures and their variation in spontaneous French. *Speech Communication*, 46(2) :119–139, 2005.

[Adda-Decker et Hallé, 2007] M. Adda-Decker et P.A. Hallé. Bayesian framework for voicing alternation and assimilation studies on large corpora in French. In *16th Congress of Phonetic Sciences*, pages 613–616, Saarbrücken, 2007.

[Adda-Decker et Lamel, 1999] M. Adda-Decker et L. Lamel. Pronunciation variants across system configuration, language and speaking style. *Speech Communication*, 29 :83–98, 1999.

[Adda-Decker, 2006] M. Adda-Decker. De la reconnaissance automatique de la parole à l'analyse linguistique de corpus oraux. In *Journées d'Étude sur la Parole*, pages 389–400, Dinard, 2006.

[Armstrong et Boughton, 1997] N. Armstrong et Z. Boughton. Identification and evaluation responses to a French accent : some results and issues of methodology. *Revue PArole*, 5–6 :27–60, 1997.

[Bartkova et Jouvet, 2004] K. Bartkova et D. Jouvet. Ensemble élargi de phonèmes pour la reconnaissance de parole avec accents. In *Colloque Modélisations pour l'IDentification des Langues*, pages 77–78, Paris, 2004.

[Bauvois, 1996] C. Bauvois. Parle-moi, et je te dirai peut-être d'où tu es. *Revue de Phonétique Appliquée*, 121 :291–309, 1996.

[Bimbot *et al.*, 2004] F. Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcìa, D. Petrovska-Delacrétaz, et D.A. Reynolds. A tutorial on text-independant speaker verification. *EURASIP journal on applied signal processing*, 4 :430–451, 2004.

[Binisti et Gasquet-Cyrus, 2003] N. Binisti et M. Gasquet-Cyrus. Les accents de marseille. *Cahiers du français contemporain*, 8 :107–129, 2003.

[Bister-Broosen, 2002] H. Bister-Broosen. Alsace. *Journal of multilingual and multicultural development*, 23 :98–111, 2002.

[Boersma et Weenink, 2008] P. Boersma et D. Weenink. *Praat : doing phonetics by computer*, 2008. www.praat.org.

[Boersma, 1993] P. Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences 17*, pages 97–110, Université d'Amsterdam, 1993.

[Borrel, 1999] A. Borrel. Les voyelles nasales du français parlé à Toulouse. *Travaux de didactique du français langue étrangère*, 41 :15–26, 1999.

[Bothorel-Witz, 2000] A. Bothorel-Witz. Les langues en Alsace. *DiversCité Langues*, 5, 2000.

[Boula de Mareüil *et al.*, 2004] P. Boula de Mareüil, B. Brahimi, et C. Gendrot. Role of segmental and suprasegmental cues in the perception of Maghrebian-accented French. In *8th International Conference on Spoken Language Processing*, pages 2885–2889, Jeju, 2004.

[Boula de Mareüil *et al.*, 2005] P. Boula de Mareüil, B. Habert, F. Bénard, M. Adda-Decker, C. Barras, G. Adda, et P. Paroubek. A quantitative study of disfluencies in French broadcast interviews. In *ISCA TR Workshop on Disfluency in Spontaneous Speech (DISS)*, pages 27–32, Aix-en-Provence, 2005.

[Boula de Mareüil *et al.*, 2008] P. Boula de Mareüil, A. Rilliard, et A. Allauzen. A diachronic study of prosody through French audio archives. In *4th Conference on Speech Prosody*, pages 531–534, Campinas, 2008.

[Boula de Mareüil *et al.*, 2009] P. Boula de Mareüil, C. Woehrling, et M. Adda-Decker. Apports du traitement automatique à une approche linguistique de la variation régionale dans la parole. In *Méthodes et outils pour l'analyse phonétique de grands corpus oraux*, éditeur N. Nguyen. Hermès, 2009. à paraître.

[Boula de Mareüil et Adda-Decker, 2002] P. Boula de Mareüil et M. Adda-Decker. Studying pronunciation variants in French by using alignment techniques. In *7th International Conference on Spoken Language Processing*, pages 2274–2277, Denver, 2002.

[Boula de Mareüil, 2007] P. Boula de Mareüil. Traitement du schwa : de la synthèse à l'alignement. In *5es Journées d'Études Linguistiques*, pages 181–189, Nantes, 2007.

[Breiman *et al.*, 1984] L. Breiman, J. H. Friedman, R. A. Olshen, et C. J. Stone. Classification and regression trees. Rapport technique, Wadsworth International, Monterey, 1984.

[Bürki *et al.*, 2008] A. Bürki, I. Racine, H. N. Andreassen, C. Fougeron, et U. H. Frauenfelder. Timbre du schwa en français et variation régionale : une étude comparative. In *Journées d'Étude sur la Parole*, pages 293–296, Avignon, 2008.

[Burger et Draxler, 1998] S. Burger et C. Draxler. Identifying dialects of German from digit strings. In *First International Conference on Language Resources and Evaluation*, pages 1253–1357, Grenade, 1998.

[Carton *et al.*, 1983] F. Carton, M. Rossi, D. Autesserre, et P. Léon. *Les Accents des Français*. Hachette, Paris, 1983.

[Carton *et al.*, 1991] F. Carton, R. Espesser, et J. Vaissière. Étude sur la perception de l'accent régional du nord et de l'est de la France. In *XIIe Congrès International des Sciences Phonétiques*, pages 422–425, Aix-en-Provence, 1991.

[Chang et Lin, 2001] C.-C. Chang et C.-J. Lin. *LIBSVM : a library for support vector machines*, 2001.

[Childers, 1978] D.G. Childers. *Modern Spectrum Analysis*. IEEE Press, 1978.

[Clairet, 2005] S. Clairet. Les voyelles nasales en français méridional et non méridional : une étude aérodynamique et acoustique. In *Langues : histoires et usages dans l'aire méditerranéenne*, éditeur T. Arnavielle, pages 239–246. l'Harmattan, Paris, 2005.

[Clopper et Pisoni, 2004] C.G. Clopper et D.B. Pisoni. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, 32 :111–140, 2004.

[Coquillon *et al.*, 2000] A. Coquillon, A. Di Cristo, et M. Pitermann. Marseillais et Toulousains gèrent-ils différemment leurs pieds ? Caractéristiques prosodiques du schwa dans les parlers méridionaux. In *Journées d'Étude sur la Parole*, pages 89–92, Aussois, 2000.

[Coquillon, 2005] A. Coquillon. *Caractérisation prosodique du parler de la région marseillaise*. Thèse de Doctorat, Université de Provence, Aix-en-Provence, 2005.

[Crystal, 2003] D. Crystal. *A dictionary of linguistics and phonetics*. Blackwell, 2003.

[D'Arcy *et al.*, 2005] S.M. D'Arcy, M.J. Russell, S.R. Browning, et M.J. Tomlinson. The Accents of the British Isles (ABI) Corpus. In *Modélisations pour l'Identification des Langues*, pages 115–119, Paris, 2005.

[Delais-Roussarie et Durand, 2003] É. Delais-Roussarie et J. Durand. *Corpus et variation en phonologie du français : méthodes et analyses*. Presses Universitaires du Mirail, Toulouse, 2003.

[Delvaux *et al.*, 2002] V. Delvaux, T. Metens, et A. Soquet. French nasal vowels : acoustic and articulatory properties. In *International Conference on Spoken Language Processing*, pages 53–56, Denver, 2002.

[Disner, 1980] S. F. Disner. Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America*, 67 :253–261, 1980.

[Durand *et al.*, 1987] J. Durand, C. Slater, et H. Wise. Observations on schwa in southern French. *Linguistics*, 25(5) :883–1004, 1987.

[Durand *et al.*, 2002] J. Durand, B. Laks, et C. Lyche. La phonologie du français contemporain : usages, variétés et structure. In *Romanistische Korpuslinguistik - Korpora und gesprochene Sprache/Romance Corpus Linguistics - Corpora and Spoken Language*, éditeurs C. Pusch et W. Raible, pages 93–106. Gunter Narr Verlag, Tübingen, 2002.

[Durand *et al.*, 2005] J. Durand, B. Laks, et C. Lyche. Un corpus numérisé pour la phonologie du français. In *La linguistique de corpus*, éditeur G. Williams, pages 205–217. Presses Universitaires de Rennes, Rennes, 2005.

[Durand et Lyche, 2004] J. Durand et C. Lyche. Structure et variation dans quelques systèmes vocaliques du français : l'enquête Phonologie du français contemporain (PFC). In *Variation et francophonie*, éditeurs A. Coveney et C. Sanders, pages 217–240. L'Harmattan, Paris, 2004.

[Durand, 1995] J. Durand. Alternances vocaliques en français du midi et phonologie du gouvernement. *Lingua*, 95 :27–50, 1995.

[Ferragne, 2008] E. Ferragne. *Étude phonétique des dialectes modernes de l'anglais des Îles Britanniques : vers l'identification automatique du dialecte*. Thèse de Doctorat, Université Lumière Lyon II, 2008.

[Fougeron et Smith, 2000] C. Fougeron et C. L. Smith. French. In *Handbook of the International Phonetic Association : A guide to the use of the International Phonetic Alphabet*, éditeur J.H. Esling, pages 78–81. Cambridge University Press, Cambridge, 2000.

[Francard, 2000] M. Francard. 'L'accent belge' : mythes et réalités. In *French accents : phonological and sociolinguistic perspectives*, éditeurs M.-A. Hintze, T. Pooley, et A. Judge, pages 251–268. AFLS/CiLT, Londres, 2000.

[Galliano *et al.*, 2006] S. Galliano, É. Geoffrois, G. Gravier, J.-F. Bonastre, D. Mostefa, et K. Choukri. Corpus description of the ESTER evaluation campaign for the rich transcritpion of French broadcast news. In *LREC*, pages 139–142, Genoa, 2006.

[Gauvain *et al.*, 2005] J.-L. Gauvain, G. Adda, M. Adda-Decker, A. Allauzen, V. Gendner, L. Lamel, et H. Schwenk. Where are we in transcribing French broadcast news ? In *9th European Conference on Speech Communication and Technology*, pages 1665–1668, Lisbonne, 2005.

[Gendrot et Adda-Decker, 2005] C. Gendrot et M. Adda-Decker. Impact of duration on F_1/F_2 formant values of oral vowels : an automatic analysis of large broadcast news corpora in French and German. In *9th European Conference on Speech Communication and Technology*, pages 2453–2456, Lisbonne, 2005.

[Ghorshi *et al.*, 2007] S. Ghorshi, S. Vaseghi, et Q. Yan. Cross-entropic comparison of the effects of accent, speaker and database recording on spectral features of English accents. In *EUSIPCO*, pages 2365–2369, Poznań, 2007.

[Gilliéron et Edmont, 1902 1910] J. Gilliéron et E. Edmont. *Atlas linguistique de la France*. Champion, Paris, 1902–1910.

[Goebl, 2002] H. Goebl. Analyse dialectométrique des structures de profondeur de l'ALF. *Revue de linguistique romane*, 66(261–262) :5–63, 2002.

[Grosjean *et al.*, 2007] F. Grosjean, S. Carrard, C. Godio, L. Grosjean, et J. Dommergues. Long and short vowels in Swiss French : their production and perception. *French language studies*, 17 :1–19, 2007.

[Hallé et Adda-Decker, 2007] P.A. Hallé et M. Adda-Decker. Voicing assimilation in journalistic speech. In *16th International Congress of Phonetic Sciences*, pages 493–496, Saarbrücken, 2007.

[Hambye et Francard, 2004] P. Hambye et M. Francard. Le français dans la communauté Wallonie-Bruxelles. Une variété en voie d'autonomisation ? *French language studies*, 14 :41–59, 2004.

[Hambye et Simon, 2004] P. Hambye et A.-C. Simon. The production of social meaning via the association of variety and style : A case study of Liège Belgian French vowel lengthening. *Canadian Journal of Linguistics*, 49 :1001–1025, 2004.

[Hansen *et al.*, 2004] J.H.L. Hansen, U. Yapanel, R. Huang, et A. Ikeno. Dialect analysis and modelling for automatic classification. In *International Conference on Spoken Language Processing*, pages 1569–1572, Jeju, 2004.

[Harrison, 2004] P. Harrison. Variability of formant measurements. Master thesis, University of York, 2004.

[Hauchecorne et Ball, 1997] F. Hauchecorne et R. Ball. L'accent du Havre : un exemple de mythe linguistique. *Langage et société*, 82 :5–25, 1997.

[Heeringa *et al.*, 2009] W. Heeringa, K. Johnson, et C. Gooskens. Mesuring Norwegian dialect distances using acoustic features. *Speech Communication*, 2009. à paraître.

[Heeringa, 2004] W. Heeringa. *Measuring dialect pronunciation differences using Levenstein distance*. Thèse de Doctorat, Rijksuniversiteit, Groningen, 2004.

[Hintze *et al.*, 2000] éditeurs M.-A. Hintze, T. Pooley, et A. Judge. *French accents : phonological and sociolinguistic perspectives*. AFLS/CiLT, Londres, 2000.

[Hsu et Lin, 2002] C.-W. Hsu et C.-J. Lin. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 13 :415–425, 2002.

[Huckvale, 2004] M. Huckvale. ACCDIST : a metric for comparing speakers' accents. In *International Conference on Spoken Language Processing*, pages 29–32, Jeju, 2004.

[Huckvale, 2007] M. Huckvale. Hierarchical clustering of speakers into accents with the ACCDIST metric. In *16th International Congress of Phonetic Sciences*, pages 1821–1824, Saarbrücken, 2007.

[Jankowski *et al.*, 1999] L. Jankowski, C. Astésano, et A. Di Cristo. The initial rhythmic accent in French : Acoustical and perceptual prosodic cues. In *14th International Congress of Phonetic Sciences*, pages 257–260, San Francisco, 1999.

[Labov, 1972] W. Labov. *Sociolinguistic patterns*. University of Pennsylvania Press, Philadelphia, 1972.

[Lamel et Gauvain, 2003] L. Lamel et J.-L. Gauvain. Speech recognition. In *The Oxford Handbook of Computational Linguistics*, éditeur R. Mitkov. Oxford University Press, New York, 2003.

[Léon, 1992] P. Léon. *Phonétisme et prononciation du français*. Armand Colin, Paris, 1992.

[Léon, 1993] P. Léon. *Précis de phonostylistique. Parole et expressivité*. Fernand Nathan, Paris, 1993.

[Léonard, 1991] J.L. Léonard. *Variation dialectale et microcosme anthropologique : l'île de Noirmoutier*. Thèse de Doctorat, Université de Provence, Aix-en-Provence, 1991.

[MacQueen, 1967] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, Berkeley, 1967.

[Mahalanobis, 1936] P. C. Mahalanobis. On the generalised distance in statistics. In *Proceedings of the National Institute of Sciences of India*, volume 2(1), pages 49–55, 1936.

[Malécot et Lindsay, 1976] A. Malécot et P. Lindsay. The neutralization of /ɛ̃/-/œ̃/ in French. *Phonetica*, 33 :45–61, 1976.

[Malderez, 1991] I. Malderez. Neutralisation des voyelles nasales chez les enfants d'Île-de-France. In *International Congress of Phonetic Sciences*, pages 174–177, Aix-en-Provence, 1991.

[Manning et Schütze, 1999] C.D. Manning et H. Schütze. *Foundations of statistical natural language processing*. The MIT Press, Cambridge, 1999.

[Martinet, 1945] A. Martinet. *La prononciation du français contemporain*. Droz, Paris, 1945.

[Martinet, 1969] A. Martinet. C'est jeuli, le Mareuc ! In *Le français sans fard*, pages 345–355. PUF, Paris, 1969.

[Mertens, 1993] P. Mertens. Accentuation, intonation et morpho-syntaxe. *Travaux de linguistique*, 26 :26–69, 1993.

[Montagu, 2004] J. Montagu. Les sons sous-jacents aux voyelles nasales en français parisien : indices perceptifs des changements. In *Journées d'Étude sur la Parole*, pages 385–388, Fès, 2004.

[Moreau et Thiam, 1995] M.-L. Moreau et N. Thiam. Comment je reconnais les variétés du wolof, le discours des adolescents sur les variétés régionales et ethniques du wolof. *Sciences et techniques du langage*, 1 :49–64, 1995.

[Métral, 1977] P. Métral. Le vocalisme du français en Suisse romande : considérations phonologiques. *Cahiers Ferdinand de Saussure*, 31 :147–176, 1977.

[Nearey, 1989] T.M. Nearey. Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, 85 :2088–2113, 1989.

[Pellegrino et André-Obrecht, 2000] F. Pellegrino et R. André-Obrecht. Automatic language identification : an alternative approach to phonetic modelling. *Signal Processing*, 80 :1231–1244, 2000.

[Peterson et Barney, 1952] G.E. Peterson et H.L. Barney. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24 :175–184, 1952.

[Pohl, 1983] J. Pohl. Quelques caractéristiques de la phonologie du français parlé en Belgique. *Langue française*, 60(1) :30–41, 1983.

[Preston, 1989] D.R. Preston. *Perceptual dialectology*. Foris Publications, Dordrecht, 1989.

[Preston, 1993] D.R. Preston. Folk dialectology. In *American Dialect Research*, éditeur D.R. Preston, pages 333–378. John Benjamins, Amsterdam/Philadelphia, 1993.

[R Development Core Team, 2008] R Development Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0, www.R-project.org.

[Rispail et Moreau, 2004] M. Rispail et M.-L. Moreau. Francique et français : l'identification des accents de part et d'autre des frontières. *Glottopol*, 4, 2004. (www.univ-rouen.fr/dyalang/glottopol/telecharger/numero_4/gpl404rispmor.pdf).

[Rouas, 2007] J.-L. Rouas. Automatic prosodic variations modelling for language and dialect discrimination. *IEEE Transactions on Audio, Speech and Language Processing*, 15 :1904–1911, 2007.

[Salvi, 2003] G. Salvi. Accent clustering in Swedish using the Bhattacharyya distance. In *15th International Congress of Phonetic Sciences*, pages 1149–1152, Barcelone, 2003.

[Schwab *et al.*, 2008] S. Schwab, I. Racine, et J.-P. Goldman. Les Vaudois parlent-ils réellement plus lentement que les Parisiens ? Un début de réponse. In *Journées Phonologie du Français Contemporain*, Paris, 2008.

[Séguy, 1973] J. Séguy. La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de linguistique romane*, 37 :1–24, 1973.

[Singy, 1995 1996] P. Singy. Les francophones de périphérie face à leur langue : étude de cas en Suisse romande. *Cahiers Ferdinand de Saussure*, 49 :213–235, 1995–1996.

[Singy, 2000] P. Singy. Les disparités entre générations en Suisse romande. In *French accents : phonological and sociolinguistic perspectives*, éditeurs M.-A. Hintze, T. Pooley, et A. Judge, pages 269–287. AFLS/CiLT, Londres, 2000.

[Sjölander, 2006] K. Sjölander. *the Snack Sound Toolkit*, 2006. www.speech.kth.se/snack/.

[Sobotta, 2006] E. Sobotta. *Phonologie et migration – Aveyronnais et Guadeloupéens à Paris*. Thèse de Doctorat, Ludwig-Maximilians-Universität München et Université Paris X, 2006.

[Syrdal et Gopal, 1986] A.K. Syrdal et H.S. Gopal. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79 :1086–1100, 1986.

[Taylor, 1996] J. Taylor. La dynamique des voyelles nasales à Aix-en-Provence. *La Linguistique*, 32 :79–90, 1996.

[ten Bosch, 2000] L. ten Bosch. ASR, dialects and acoustic/phonological distances. In *International Conference on Spoken Language Processing*, pages 1009–1012, Beijing, 2000.

[' t Hart *et al.*, 1991] J. ' t Hart, R. Collier, et A. Cohen. *A perceptual study of intonation : an experimental-phonetic approach to speech melody*. CUP, Cambridge, 1991.

[Thomas, 1991] A. Thomas. Évolution de l'accent méridional en français niçois : les nasales. In *XXI^e Congrès International des Sciences Phonétiques*, pages 194–197, Aix-en-Provence, 1991.

[Vajta, 2002] K. Vajta. Le français en Alsace : premiers résultats d'une enquête. *Romansk Forum*, 16 :823–834, 2002.

[van Bezooijen et Gooskens, 1999] R. van Bezooijen et C. Gooskens. Identification of Language Varieties. Contribution of Different Linguistic Levels. *Journal of Language and Social Psychology*, 18(1) :31–48, 1999.

[Van Compernolle, 2001] D. Van Compernolle. Recognizing speech of goats, wolves, sheep and... non-natives. *Speech Communication*, 35(1–2) :71–79, 2001.

[Vieru-Dimulescu et Boula de Mareüil, 2004] B. Vieru-Dimulescu et P. Boula de Mareüil. Contribution de la prosodie à la perception de l'accent étranger : une étude à base de parole naturelle italienne/espagnole modifiée. In *Colloque Modélisations pour l'IDentification des Langues*, pages 139–144, Paris, 2004.

[Vieru-Dimulescu, 2008] B. Vieru-Dimulescu. *Caractérisation et identification d'accents étrangers en français*. Thèse de Doctorat, Université Paris XI, 2008.

[Walter, 1982] H. Walter. *Enquête phonologique et variétés régionales du français*. Presses Universitaires de France, Paris, 1982.

[Walter, 1988] H. Walter. *Le français dans tous les sens*. Robert Laffont, Paris, 1988.

[Welby, 2007] P. Welby. The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication*, 49 :28–48, 2007.

[Williams *et al.*, 1999] A. Williams, P. Garrett, et N. Coupland. Dialect recognition. In *Handbook of Perceptual Dialectology*, éditeur D.R. Preston, pages 345–358. John Benjamins, Amsterdam/Philadelphia, 1999.

[Yan et Vaseghi, 2002] Q. Yan et S. Vaseghi. A Comparative Analysis of UK and US English Accents in Recognition and Synthesis. In *IEEE ICASSP*, pages 413–416, Orlando, 2002.