Éléments de phonétique acoustique

Yohann Meynadier

Chapitre 1

Éléments de phonétique acoustique

Yohann MEYNADIER

1.1. Introduction

Les grands corpus oraux constituent des ressources scientifiques essentielles et de plus en plus centrales dans l'élaboration des théories sur la parole et le langage, à tous les niveaux du fonctionnement et de la structuration linguistiques et du comportement verbal : pragmatique, sémantique, syntaxique, morphologique, phonologique, phonétique, etc. Dans tous ces domaines, l'automatisation de procédures de segmentation, d'étiquetage, d'annotation, d'extraction de paramètres, de prise de mesures, de collecte, de traitement et d'analyse de données précises et variées est un enjeu fondamental pour l'assise de modèles et d'hypothèses linguistiques sur des descriptions factuelles et empiriques à grande échelle, notamment dans une approche écologique de la parole. Dans cet enjeu, la phonétique acoustique occupe une place décisive, puisque tout niveau linguistique d'analyse (semi-)automatique de grands corpus oraux implique ou présuppose une lecture, un traitement ou une analyse du signal acoustique de paroles enregistrées. Les connaissances acoustiques sont donc indispensables, non seulement pour les études phonétiques, mais bien au-delà, pour toute étude de larges corpus de parole, même si la matière sonore n'est pas précisément l'objet linguistique investigué.

Avant d'aborder les traitements spécifiques appliqués pour l'analyse phonétique de grands corpus oraux, il convient d'introduire l'acoustique et la phonétique acoustique, segmentale et suprasegmentale. Ainsi, ce chapitre expose les notions et paramètres de base nécessaires à l'analyse et la description acoustiques des sons de la parole, ainsi que les principales propriétés acoustiques des voyelles et des consonnes du français, en lien avec leur articulation et leur coarticulation dans la parole spontanée.

1.2. Acoustique générale

L'acoustique de la parole se réfère à la phase de transmission du son sous la forme d'une onde sonore dans l'air ambiant où baignent généralement les locuteurs. Un son est physiquement un phénomène de variations répétées plus ou moins régulières de pression qui se propage à une certaine vitesse dans un milieu élastique. La vitesse de propagation du son dépend de la densité entre les particules qui le composent. Plus cette densité est importante, plus le son se propage rapidement, par exemple à environ 5200 m/s dans l'acier, 1480 m/s dans l'eau et 340 m/s dans l'air. Les fluctuations de pression sont générées par la vibration d'un corps : un moteur en marche, une cymbale percutée, les

feuilles des arbres sous l'effet du vent, ou encore les cordes vocales lors de l'émission de parole. Dans l'atmosphère, quand un corps vibre, il impose directement un mouvement réitéré d'oscillations aux particules d'air qui l'entourent. Ces oscillations créent des micro-vagues de surpression et de sous-pression de l'atmosphère qui se diffusent de proche en proche dans l'espace au cours du temps et qui composent l'onde sonore. A l'inverse du vent, l'onde acoustique n'est donc pas un déplacement de l'air, mais un mouvement répété qui se diffuse dans l'air à la manière du mouvement que l'on perçoit lors de la chute en série de dominos agencés en circuit. Quand ces oscillations de la pression atmosphérique sont trop lentes (moins de 16 par seconde) ou trop rapides (plus de 20 000 par seconde), elles constituent des sons inaudibles pour l'oreille humaine : respectivement des infrasons et des ultrasons. Ils peuvent néanmoins être captés par d'autres types de récepteur, selon leurs caractéristiques de réponse au signal acoustique, comme le système auditif d'autres espèces animales ou bien un microphone.

Différents types de sons sont distingués en fonction de caractéristiques physiques qui définissent l'onde acoustique : principalement leur périodicité (en lien avec la fréquence) et leur complexité. Une première catégorie de sons concerne des ondes périodiques simples, appelées communément 'sons purs', par exemple la note musicale de référence *la* produite par un diapason. Ces sons sont caractérisés par un cycle d'oscillation simple, régulier et répété dans le temps, correspondant à une vague de surpression atmosphérique suivie d'une vague symétrique de sous-pression. Ils sont aisément identifiables sous la forme d'une simple onde acoustique sinusoïdale sur un oscillogramme, comme illustrée dans la figure 1.1.a. Sur un mode non sonore, le type de mouvement de la pression de l'air impliqué dans ces sons est comparable au balancement libre d'un pendule autour de sa position d'immobilité. Chaque cycle de balancement symétrique autour de cette position se répète à intervalle identique et dure donc le même temps. La durée de ce cycle parfaitement régulier dans le temps est appelée 'période' (T). Un son périodique simple est donc déterminé par une période unique, à partir de laquelle est calculée la fréquence du son. La fréquence (F) se comprend comme le nombre de cycles périodiques de l'onde acoustique décomptés en une seconde (s.). Mathématiquement, elle est définie par l'inverse de la période (F = 1/T), et s'exprime en hertz (Hz). Ainsi, si un son pur a une période de cycle de 0,02 s. – soit 20 millisecondes (ms) – sa fréquence est alors de 50 Hz – à savoir 1 s. divisée par 0,02 s. – c'est-à-dire le nombre de fois que le cycle de 20 ms se répète par seconde. Un son périodique simple se caractérise donc aussi par une fréquence unique. Plus la fréquence d'un son est basse, plus il est perçu comme grave ; plus sa fréquence est élevée, plus il est perçu comme aigu (dans la limite des seuils et différentiels minimaux du système de traitement auditif[1]). Si la fréquence d'un même son pur est constante, son amplitude, autre dimension fondamentale de l'onde acoustique, peut évoluer indépendamment au cours du temps. L'amplitude acoustique correspond à l'écart mesuré entre la variation maximale de pression d'air atteinte lors du son et la pression atmosphérique de référence. Si nombre d'échelles de mesure de cette dimension acoustique peuvent être utilisées, l'hectopascal (hPa) est la plus naturelle en tant qu'unité de la pression. Néanmoins le décibel (dB) est l'unité la plus couramment

[1] Voir notamment Moore (1989).

employée, essentiellement pour des raisons pratiques de calcul et de rapprochement avec la réponse sensorielle non linéaire de l'oreille humaine[2]. La figure 1.1a présente un son périodique simple amorti, c'est-à-dire dont la vibration faiblie graduellement en amplitude au cours du temps (et dans l'espace), du fait de la perte d'énergie du mouvement vibratoire due notamment aux forces de frottement, d'inertie, etc., jusqu'à son arrêt définitif, à la façon dont une balle s'arrête librement de rebondir. Alors que l'amplitude de la vibration diminue progressivement, la fréquence de cette vibration sonore reste stable tout au long du son. Transposé à l'image du pendule, ce dernier oscille de moins en moins fort, mais avec un même rythme régulier de balancement jusqu'à s'immobiliser dans sa position d'équilibre. Amplitude et fréquence (ou période) du son constituent donc des paramètres acoustiques indépendants. Reste que les ondes périodiques simples correspondent à des sons totalement absents de la nature, et donc de la parole. Les sons naturels sont toujours acoustiquement plus complexes. Leur complexité peut être régulière comme pour les sons périodiques complexes, appelés aussi 'sons harmoniques', tels que les notes d'instruments musicaux ou les voyelles de la parole, ou bien irrégulière comme pour les sons apériodiques, plus connus sous le terme de 'bruit', tels que les explosions, les battements, les frottements, les chuintements, qu'on peut trouver par ailleurs dans certaines consonnes de la parole.

Les sons périodiques complexes (figure 1.1c) proviennent de la vibration composée d'un corps, mais identique dans le temps. L'exemple de la vibration d'une corde tendue illustre parfaitement les oscillations de pression d'air caractéristiques de ce type de sons. La vibration d'une corde répond à différents registres ou modes vibratoires parfaitement réguliers, simultanés et associés. Communiquée à l'atmosphère, cette vibration composite de la corde se propage sous la forme d'une onde périodique complexe constituée d'un ensemble d'ondes périodiques simples combinées, appelées 'partiels' ou 'harmoniques'. Les harmoniques d'un son présentent des relations systématiques et directes de fréquence et d'amplitude. Le théorème de Fourier, appliqué à l'analyse acoustique (par la transformée rapide de Fourier ou FFT), explicite mathématiquement ces liens, tels que les harmoniques sont des multiples entiers de l'harmonique présentant la période la plus longue, c'est-à-dire la fréquence la plus basse. Cet harmonique le plus grave en déterminant la propre fréquence d'oscillation d'un son périodique complexe est donc essentiel, d'où sa dénomination de 'fréquence fondamentale' (ou F_0), les autres composantes harmoniques lui étant assujetties : la fréquence du 2nd harmonique est égale à $2\times F_0$, du 3ème à $3\times F_0$, … du 7ème à $7\times F_0$, etc. L'analyse FFT (figure 1.2) permet ainsi la décomposition fréquentielle des sons périodiques complexes. Sa représentation graphique, appelée 'spectre', présente en abscisse la fréquence et en ordonnée l'amplitude de chacun des harmoniques du son[3]. L'oscillogramme représente la forme de l'onde acoustique du son, c'est-à-dire la variation d'amplitude (en ordonnée) de la pression de l'air au cours du temps (en abscisse). La forme de l'onde acoustique d'un son périodique complexe manifeste des relations d'amplitude avec ses différents

[2] Nous laisserons cependant de côté ces questions d'échelle plus secondaires à notre propos. Pour plus de détails, voir notamment Martin (2008).

[3] On notera qu'un son périodique simple ne présente qu'une seule composante spectrale, puisqu'il est caractérisé par une fréquence unique (la F_0), relatif à son mode vibratoire singulier (figure 1.1b).

composants fréquentiels, de telle sorte qu'elle est le résultat de la somme, point par point temporel, de l'amplitude de chacun d'eux (figure 1.2).

Les sons apériodiques (figure 1.1e-g), ou bruits, résultent de la vibration chaotique d'un corps. Du point de vue aérodynamique, ils se caractérisent par une mise en turbulence de l'air engendrant des fluctuations irrégulières de la pression atmosphérique. L'onde sonore complexe consécutive ne présente aucune constante ni de temps, ni de forme, ni de relations entre ses composantes. Ce sont des sons sans période d'oscillation, et donc sans F_0 ni harmoniques. Les fréquences qui les composent et leur amplitude sont totalement aléatoires et indépendantes. Comparés à un son harmonique, spectre FFT et oscillogramme de sons apériodiques rendent bien compte de leur composition fréquentielle complexe et chaotique. Deux types de sons apériodiques sont distingués sur la seule base de leur durée. Les bruits continus ou longs (figure 1.1e) durent plusieurs dizaines de millisecondes, tels que les bruits de grincement, chuintement, frottement ou friction, comme ceux impliqués dans les consonnes de la parole appelées 'fricatives'. Les bruits impulsionnels ou brefs (figure 1.1g) ne dépassent guère la dizaine de millisecondes, tels que les bruits de battement, claquement, implosion ou explosion, comme ceux impliqués dans les consonnes de la parole appelées 'plosives'.

Jusque là, les sons ont été considérés uniquement comme des événements vibratoires isolés ignorant le phénomène essentiel de résonance acoustique. Or, les vibrations sonores n'apparaissent pas dans un environnement vide. Dans le monde réel, les sons entrent en résonance ou sont mis en résonance par propagation à d'autres corps voisins. Ainsi, les vibrations d'un corps peuvent se transmettre à un autre corps assez proche. Par exemple, le bruit du moteur d'un bus peut mettre en vibration les vitres de fenêtres d'immeubles aux alentours. De la même manière, le son du tic-tac d'une montre sera plus fort, si celle-ci est posée dans une boîte ouverte que tenue à la main. En effet, les faibles vibrations, à peine audibles, de la montre se transmettent aux parois de la boîte. Les vibrations consécutives de la boîte affectent alors un bien plus grand volume d'air que ne le feraient à elles seules les vibrations de la montre. Le son de la montre résonne dans la boîte. La résonance acoustique est de ce fait un phénomène d'amplification vibratoire de l'air. Mais cette amplification n'affecte pas de manière homogène toutes les composantes fréquentielles d'un son, c'est-à-dire les différents harmoniques d'un son périodique ou les multiples fréquences d'un bruit. Chaque corps selon sa taille, sa forme, son volume ou sa matière ne vibre pas de la même façon, et donc ne répond pas identiquement à l'excitation provoquée par une source sonore. On parle de 'fréquences naturelles de résonance' propre à chaque corps pour désigner le fait que celui-ci est plus facilement mis en vibration par des sons composés de fréquences proches ou égales aux siennes et beaucoup plus difficilement ou pas du tout à d'autres plus éloignées. Ainsi, si la montre est posée dans une boîte différente, son tic-tac ne résonnera pas exactement de la même manière et sera perçu avec un timbre différent, car cette nouvelle boîte a d'autres fréquences naturelles de résonance. Dans une grande boîte son tic-tac est plus grave que dans une plus petite, car les fréquences naturelles de résonance d'une petite boîte favorisent des composantes plus aigües du son, alors que celles d'une grande boîte favorisent des fréquences plus graves. Cela signifie que la résonance modifie exclusivement l'amplitude des composantes fréquentielles d'un son selon les

caractéristiques de réponse vibratoire propres au résonateur. De ce fait, un résonateur se comporte comme un filtre acoustique : laissant passer, renforçant, atténuant ou bloquant certaines fréquences du son complexe. Il modifie la répartition de l'énergie acoustique dans les différentes fréquences du spectre du son, appelée 'balance spectrale', terme correspondant sur le mode perceptif au timbre sonore. La résonance acoustique est donc un phénomène d'amplification fréquentielle sélective, fondamentale notamment s'agissant des sons linguistiques[4].

1.3. Les sons de la parole

La théorie *source-filtre*, développée par Fant (1960), constitue le modèle de référence pour la production des sons de la parole. Il assimile le système de production acoustique de la parole à un système combinant un ou deux générateurs d'un son de base, appelés 'source' à un ensemble de deux à quatre résonateurs couplés, appelés 'filtre', modulant en sortie certaines propriétés du son source.

Dans ce cadre, les sons complexes, décrits précédemment en dehors du phénomène de résonance, définissent exclusivement la source acoustique des sons de parole. Ainsi, la source peut être de types différents et combinables dans la parole. D'une part, la vibration des cordes vocales constitue une source harmonique générant un son quasi périodique complexe : l'onde glottique, appelée communément 'voix' ou 'voisement'. Ce son est à l'origine des sons voisés de la parole, tels que les voyelles et un grand nombre de consonnes, dont [m l ʁ z g]. D'autre part, des constrictions, opérées par le rapprochement étroit de différents articulateurs tout au long du conduit vocal, de la glotte aux lèvres, peuvent également générer, sous la forme d'un bruit, une source apériodique indépendante de la première. Cette apériodicité est le fait d'un échappement de l'air phonatoire hors du conduit vocal rendu difficile par une constriction dite 'critique', c'est-à-dire trop étroite pour permettre un flux d'air sortant laminaire, à savoir régulier, fluide et silencieux. Ainsi, un mince canal d'écoulement produit un flux d'air sortant turbulent, à savoir chaotique et bruyant, du fait du frottement important de l'air sur les parois du conduit déviant les particules d'air de leur trajectoire parallèle initiale. Ce bruit peut, dans certains cas, se substituer totalement à la source périodique laryngée. Il peut être émis au niveau du larynx par une constriction glottique comme dans la parole chuchotée, ou au-dessus par une constriction supraglottique en différents points du conduit vocal comme pour les consonnes fricatives non voisées (ou sourdes) [f s ʃ]. Dans les cas des consonnes fricatives voisées [v z ʒ ʁ], cette source supraglottique apériodique se superpose à la source périodique glottique du voisement. Lorsque le conduit vocal est totalement fermé, l'air accumulé temporairement en amont de l'occlusion est brutalement libéré lors du relâchement de cette fermeture, ce qui crée un flux turbulent bref et soudain générant un bruit transitoire d'explosion. Dans le cas des consonnes plosives (ou occlusives) voisées [b d g], cette brusque source bruitée est produite en fin de tenue voisée (source périodique) de la consonne. Dans le cas des consonnes plosives sourdes [p t k], ce bruit bref constitue la seule source acoustique, la

[4] Voir Ghio & Pinto (2007) pour plus de détails sur la définition physique de la résonance.

tenue consonantique étant silencieuse, du fait d'une position très écartée des cordes vocales empêchant leur mise en vibration ou l'établissement d'une turbulence aérodynamique au niveau de la glotte.

Quel que soit le type de source, avant sa propagation dans l'atmosphère, le son de base émis est modulé par le filtre acoustique que constitue le conduit vocal. Le conduit vocal forme une sorte de tube coudé composé de la cavité pharyngale (partie supraglottique verticale postérieure située dans la gorge) débouchant, d'une part, sur la cavité buccale (partie supraglottique horizontale antérieure située dans la bouche), et, d'autre part, sur la cavité nasale (partie correspondant aux fosses nasales séparées de la bouche par l'os du palais dur) quand leur communication est rendue possible par l'abaissement du voile du palais (muscle prolongeant le palais dur) ouvrant le passage naso-pharyngal. La protusion des lèvres permet en bout du conduit oral de créer une sorte de quatrième cavité supplémentaire : la cavité labiale, impliquée pour dans la production des sons arrondis, comme [y u o], voire [ʃ ʒ] (Toda 2009). Chacune de ces cavités sont excitées par certaines fréquences composant le ton laryngé se propageant dans le conduit vocal, devenant ainsi de véritables résonateurs acoustiques, à la manière de la boîte mise en vibration par la montre. En réponse à cette excitation, ces résonateurs amplifient certains paquets de fréquences du son à proximité de leurs propres fréquences naturelles de résonance. Cet ensemble de résonateurs couplés se comporte alors comme un filtre acoustique complexe, en modifiant la balance spectrale du son source, lui conférant ainsi son timbre particulier.

Le modèle de conduit vocal à quatre tubes de Fant permet d'expliciter les liens complexes mais primordiaux entre les différentes configurations du conduit vocal et leur résultat acoustique en termes de profil de résonance lors de la production des sons de la parole. Le profil de résonance du conduit est mathématiquement déterminé par une fonction de transfert (figure 1.3) représentant les relations entre l'entrée (la source acoustique) et la sortie (le son de parole) du système (le conduit vocal). Outre les différences liées à la source, chaque son de la parole se caractérise par un positionnement spécifique des différents articulateurs mobiles composant le conduit vocal : principalement les lèvres, les différentes parties de la langue (de l'avant à la base : apex ou pointe, lame, dos et racine), la mandibule inférieure, le voile du palais (ou palais mou ou velum). De ce fait, chaque articulation d'un son a pour effet de déformer et sectionner différemment le conduit vocal, qui se définit alors fondamentalement par la géométrie des différentes cavités supraglottiques connectées. Selon leur taille, différentes composantes spectrales du son source sont amplifiées, donnant lieu à l'émergence de zones fréquentielles plus intenses que d'autres, correspondant aux fréquences naturelles de résonance du conduit vocal dans la configuration articulatoire adoptée pour la production du son articulé. Ces bandes fréquentielles renforcées portent le nom de 'pôles', et plus spécifiquement de 'formants' s'agissant des voyelles et des consonnes dites 'sonantes' telles que [m n l j]. La figure 1.3 explicite l'émergence des trois premiers formants de voyelles, visualisables sous la forme de maxima spectraux, selon les principales étapes du modèle *source-filtre*. Outre ces amplifications locales, certains sons de parole présentent également une atténuation importante de l'amplitude dans certaines bandes fréquentielles de leur spectre, apparaissant sous la forme de creux

locaux d'énergie. Ces creux spectraux, appelés 'zéro' ou 'antiformant', relèvent d'un phénomène d'antirésonance qui opère lorsqu'un embranchement est réalisé dans le conduit vocal, et que l'air expiré en résonance emprunte plus d'un seul canal. Cela est typiquement le cas quand la cavité nasale est couplée au conduit oral par l'abaissement du voile du palais. Conduits oral et nasal résonnent alors en parallèle, donnant lieu à des interactions complexes où certaines fréquences de résonance d'un conduit entrent en opposition ou en décalage de phase avec celles de l'autre conduit. Pour illustrer ce phénomène, imaginons un enfant bercé à un rythme régulier. Pour maintenir ou amplifier son bercement, il faut synchroniser précisément chacune des impulsions sur le berceau toujours au même moment du cycle de balancement, dans ce cas quand la nacelle atteint le point de retour maximal. Si les impulsions, bien qu'au même rythme que ce balancement, ont lieu avant ou après cet instant donné, l'enfant est bercé de moins en moins fort, car le cycle des impulsions est déphasé (non synchrone) par rapport au cycle du balancement : celui-ci est alors atténué. Si ces impulsions sont parfaitement symétriquement opposées au cycle de bercement, le balancement de la nacelle s'arrête : celui-ci est alors annulé. Le phénomène d'antirésonance acoustique répond au même principe concernant les composantes fréquentielles des sons. Les antiformants sont caractéristiques des consonnes et des voyelles dites 'nasales' comme [ɛ̃ ɑ̃ m n]. Ce phénomène apparaît également lors de la production des consonnes dites 'latérales', comme [l] en français, où la cavité orale présente une bifurcation latérale assimilable à un canal oral parallèle (Fant 1960, Johnson 1997). De façon différente, les antiformants apparaissent aussi lorsqu'une source acoustique (bruit), comme pour les consonnes plosives et fricatives, est produite dans le conduit vocal (des lèvres au pharynx) et non en son commencement au niveau du larynx. Dans ce cas l'air en résonance peut s'échapper selon deux voies vers l'avant dans la cavité antérieure, la principale, et vers l'arrière dans la cavité postérieure, plus secondaire, provoquant des effets acoustiques de couplage complexes (Stevens 1998, Harrington & Cassidy 1999). Ainsi, formants/pôles (pics spectraux) et antiformants/zéros (creux spectraux) déterminent le timbre spécifique de chaque son de la parole. Associés à la présence ou l'absence de voisement, ils constituent des caractéristiques acoustiques discriminantes pour l'identité phonétique des sons, sur lesquelles s'ancrent les propriétés phonologiques essentielles à la structuration des formes et des systèmes sonores des langues.

La plupart du temps, l'analyse phonétique s'attache à étudier ces patrons de pics spectraux des sons, bien plus que leur structure strictement harmonique, à des fins de description linguistique. Deux techniques communes d'analyse acoustique sont particulièrement dédiées à cet examen : le spectre LPC pour '*Linear Predictive Coding*' (méthode des 'coefficients de prédiction linéaire' ou de Prony) et le spectrogramme à bande large. Sans entrer dans les détails trop techniques de leur procédure[5], ces analyses permettent l'extraction des caractéristiques phonétiques des sons de la parole en termes de formants/pics, à savoir de zones d'émergence fréquentielles. Le spectre LPC permet, sur la base d'une application du modèle *source-filtre* et de méthodes d'auto-corrélation, de dégager la fonction de transfert correspondant au son analysé, et par là-même de

[5] Pour plus de détails, voir Auran et al. (ce volume), Tubach (1989), Ladefoged (1996), Johnson (1997), Harrington & Cassidy (1999), Martin (2008), notamment.

déterminer la localisation et la largeur de bande des formants, numérotés de F_1 à F_n selon un ordre croissant fréquentiellement, visualisés par des pics spectraux (figure 1.3). Cette technique est cependant moins bien adaptée à l'analyse de sons dont la fonction de transfert est plus complexe, comme les nasales ou les fricatives. Un spectre LPC, comme FFT, reste cependant une analyse acoustique statique 2D (fréquence/amplitude), telle une photographie à un instant choisi d'un son de la parole. En comparaison, un spectrogramme à bande large permet l'analyse de l'évolution des pôles fréquentiels au cours du temps, proposant ainsi une sorte de film acoustique 3D (temps/fréquence/amplitude) de la parole en continu. Le spectrogramme à bande large est conçu par la juxtaposition de spectres FFT effectués avec une petite fenêtre d'analyse (à savoir une courte portion de signal, de l'ordre d'une dizaine de ms ou moins) avec un important recouvrement temporel, c'est-à-dire décalée à petits pas (de l'ordre de quelques ms) le long du signal. Cette méthode garantie une grande précision temporelle, permettant une segmentation plus précise des sons, pour une résolution en fréquences moins fine, mais suffisante pour dégager les formants ou pôles fréquentiels, visualisables sous la forme de larges bandes horizontales plus foncées plus ils sont intenses (figure 1.3). Cette représentation acoustique dynamique constitue l'outil privilégié de l'étude phonétique de la parole continue, notamment dans les procédures de segmentation[6], d'étiquetage et de mesure automatiques du signal de parole.

1.3.1 *Les voyelles du français*

En dehors de cas de coarticulation (cf. ci-après) et de la phonation chuchotée, les voyelles sont des sons voisés. Acoustiquement, ce sont donc des sons périodiques complexes dont la résonance du voisement dans le conduit vocal produit l'émergence de formants caractéristiques. Articulatoirement, elles se réalisent principalement par un positionnement de la langue, pouvant indépendamment se coupler à une projection des lèvres (protusion), réalisant des voyelles arrondies : [y ø œ u o ɔ], et/ou à un abaissement du voile du palais, produisant des voyelles dites 'nasales' : [ɛ̃ œ̃ ɔ̃ ɑ̃]. Pour les voyelles non nasales (c'est-à-dire orales) et non arrondies [i e ɛ a] (toutes antérieures), la résonance du conduit vocal est donc essentiellement déterminée par une configuration linguale distincte pour chaque voyelle. La langue réalise dans le conduit oral une constriction plus ou moins large selon son degré d'élévation dans la cavité buccale. Le point de rapprochement maximal entre la langue et le plafond de la bouche (palais dur et mou) détermine le lieu d'articulation (Ladefoged & Maddieson 1996) : antérieur (palatal) pour [i e ɛ a y ø œ ɛ̃ œ̃] et postérieur (vélaire) pour [u o ɔ ɔ̃ ɑ̃]. D'un côté, cette constriction ne proposant pas d'obstacle important à l'échappement du flux d'air est suffisamment ouverte pour garantir un son sans bruit supraglottique qui caractériserait des sons consonantiques. D'un autre côté, cette constriction reste assez resserrée pour provoquer une segmentation du conduit vocal en cavités de résonance distinctes bien qu'en communication : la cavité avant ou buccale en aval de la constriction, et, la cavité

[6] Pour une présentation détaillée de la méthode manuelle et des indices acoustiques de segmentation spectrographique (standardisée) des sons de la parole continue, voir Machač & Skarnitzl (2009).

arrière ou pharyngale en amont. Soulignons que les différences de taille globale du conduit impliquent des variations de fréquence des formants entre individus et notamment entre sexes. Le conduit vocal féminin (≈ 15 cm) est d'environ 10-15% plus court que celui d'un homme (≈ 17 cm), particulièrement du fait de la position plus basse du larynx et de la corpulence générale plus importante de l'homme. Cette différence de taille du conduit explique en grande partie des formants vocaliques plus élevés pour une voix de femme que d'homme.

Dans le cadre de la théorie *source-filtre*, le conduit vocal lors de l'articulation des voyelles orales antérieures non arrondies [i e ɛ a] s'apparente à un système à deux tubes connectés dont la résonance respective dépend essentiellement de leur volume (ou longueur) individuel. Or, les liens entre pôles de résonance et longueur d'un tube, bien connus et formulés mathématiquement, rendent compte d'une relation inverse : plus un tube est petit (court et/ou étroit) plus ses fréquences naturelles de résonance sont hautes ; plus il est grand (long et/ou large), plus elles sont basses. Il faut aussi noter ici que pour un conduit d'une taille donnée comportant deux tubes dans leur prolongement, il s'établit une relation inverse entre la taille de chacun d'eux, et donc entre leurs pôles respectifs de résonance. Ce système à deux tubes permet de modéliser adéquatement la structure formantique des voyelles lorsqu'une simple constriction est opérée dans le conduit oral, comme cela est le cas des voyelles [i e ɛ a]. Un petit résonateur avant implique ainsi un grand résonateur arrière, et inversement (figure 1.3). Dès lors, toutes choses égales par ailleurs, plus la fréquence du formant vocalique issu de la résonance de la cavité avant est élevée, plus la fréquence du formant lié à la résonance de la cavité arrière est basse, et vice-versa. Dans la figure 1.4 (en bas à gauche), les voyelles orales du français sont localisées dans un espace acoustique représentant les variations de fréquence de leurs deux premiers formants, appelé 'plan F_1-F_2'. Les voyelles [i e ɛ a], présentant une constriction antérieure (au niveau du palais dur), diffèrent principalement entre elles par le degré de cette constriction (appelé 'aperture') selon l'abaissement graduel de la langue croissant de [i] à [a] en passant par [e] puis [ɛ]. Il est à noter ici que l'abaissement graduel de la langue s'accompagne mécaniquement d'un recul du lieu d'articulation dans le conduit plus une voyelle antérieure est ouverte, du fait de l'abaissement concomitant de la mandibule inférieure à laquelle la langue est anatomiquement attachée. Ce mouvement d'ouverture buccale a ainsi pour conséquence de réduire la taille de la cavité pharyngale et d'augmenter celle de la cavité buccale – comme les configurations articulatoires du conduit vocal de [i] et de [a] l'illustrent dans la figure 1.3. [i] est la voyelle la plus antérieure et la plus fermée. Elle présente ainsi la plus petite cavité avant et donc inversement une très grande cavité arrière, ce qui se traduit acoustiquement par un F_2 très haut (≈ 2050 Hz) et un F_1 très bas (≈ 300 Hz). Au contraire [a], étant la voyelle d'avant la moins antérieure et la plus ouverte, a une cavité avant très grande associée à un F_2 assez bas (≈ 1250 Hz) et une très petite cavité arrière affiliée à un F_1 très haut (≈ 700 Hz). [e] et [ɛ], présentant un abaissement et un avancement de la langue intermédiaires, montrent des valeurs formantiques médianes : respectivement, pour F_1 ≈ 350 Hz et ≈ 550 Hz, et pour F_2 ≈ 1950 Hz et ≈ 1700 Hz. [e] plus fermée et antérieure a un F_1 plus bas et un F_2 plus haut que [ɛ] plus ouverte et moins antérieure. Si le système à deux tubes du conduit vocal permet assez simplement de

calculer la valeur fréquentielle des formants de ces voyelles à partir de la taille des résonateurs (tubes), il doit être complexifié en un système à quatre tubes pour modéliser la structure acoustique des autres voyelles orales du français qui sont toutes arrondies. [y ø œ u o ɔ] sont en effet réalisées par une double constriction du conduit vocal : l'une linguale en avant ou en arrière dans le conduit, combinée à une seconde en bout du conduit oral produite par la protusion des lèvres. Si la prédiction mathématique des fréquences formantiques associées à ces configurations géométriques plus complexes du conduit vocal modélisé en termes de tubes est connue, elle présente néanmoins des complications importantes dépassant les limites de ce chapitre introductif[7].

A ce stade, une inspection plus empirique des fréquences représentatives des trois premiers formants relevées dans la production de 10 locuteurs français suffit à saisir les liens généraux entre articulation et acoustique des voyelles orales. La figure 1.4 (en bas à droite) présente le triangle vocalique classique de l'Alphabet Phonétique International pour les voyelles orales du français (Fougeron & Smith 1999). Cette représentation schématique d'un espace d'articulation distribue les voyelles selon leurs propriétés articulatoires distinctives : sur l'axe vertical leur aperture, sur l'axe horizontal leur lieu d'articulation combiné au degré d'arrondissement des lèvres. La voyelle non arrondie la plus fermée et antérieure [i] est placée en haut à l'extrême gauche du graphique ; la plus fermée, postérieure et arrondie [u] est située en haut à l'extrême droite ; la voyelle la plus ouverte, plutôt centrale et non arrondie [a] se situe en bas proche du centre. Les voyelles antérieures arrondies [y ø œ] sont reportées dans une position décalée vers la droite par rapport à leurs pendants non arrondis [i e ɛ] ayant la même configuration linguale. Comparer cette distribution articulatoire des voyelles aux variations de leur F_1, F_2 et F_3 permet d'appréhender globalement les relations entre propriétés articulatoires et structure formantique des voyelles, illustrée par leur position dans les plans F_1-F_2 et F_2-F_3 de la figure 1.4. Communément, F_1 est associé à la résonance de la cavité arrière. Or, la taille de cette cavité est plus particulièrement sensible au degré d'abaissement de la langue et de la mandibule, c'est-à-dire à l'aperture des voyelles. Plus elles sont ouvertes plus la cavité pharyngale est petite et donc sa fréquence de résonance naturelle élevée, d'où un F_1 régulièrement plus haut des voyelles fermées à ouvertes, qu'elles soient antérieures non arrondies : [i] < [e] < [ɛ] < [a], antérieures arrondies : [y] < [ø] < [œ], ou postérieures (arrondies) : [u] < [o] < [ɔ]. Quel que soit leur lieu d'articulation ou leur arrondissement, F_1 aura une fréquence sensiblement similaire pour chaque série de voyelles de même aperture : fermées [i y u] < mi-fermées [e ø o] < mi-ouvertes [ɛ œ ɔ] < ouverte [a]. Principalement, F_2 est rattaché à la résonance de la cavité avant, plus drastiquement affectée par le lieu d'articulation linguale. Plus les voyelles sont antérieures, plus la cavité avant est petite, et donc plus la fréquence de F_2 augmente. Ainsi, à protusion labiale et aperture équivalentes, le F_2 de [y] est nettement supérieur à celui de [u], celui de [ø] à celui de [o], et de [œ] à [ɔ]. Comme cela a été exposé précédemment, dans une moindre mesure le F_2 des voyelles antérieures (arrondies [y ø

[7] Pour une présentation plus complète du modèle à tubes du conduit vocal sortant de l'objet de cette introduction, se reporter notamment à Fant (1960), Johnson (1997), Stevens (1998), Harrington & Cassidy (1999).

œ] ou non [i e ɛ a]) diminue avec l'aperture du fait de l'effet de l'abaissement mandibulaire sur l'antériorité de la constriction linguale. Concernant les voyelles postérieures [u o ɔ], le phénomène est inversé. F_2 augmente avec l'abaissement lingual et mandibulaire. Dans ce cas, la constriction se fait avec une partie plus postérieure de la langue massée à l'arrière de la cavité buccale. Dans cette configuration linguale, un abaissement de la mandibule provoque un avancement du point de constriction maximale, et donc de son lieu d'articulation. Mais F_2 est également sensible au geste des lèvres. L'arrondissement labial crée une sorte de cavité supplémentaire en bout du conduit vocal qui peut en partie s'analyser comme une extension de la cavité en avant de la constriction linguale. Dès lors, lorsque qu'une voyelle produite avec une même aperture à un même lieu d'articulation est arrondie, la cavité buccale s'en trouve allongée, et en conséquence F_2 est abaissée. Ainsi, [y] montre un F_2 plus bas que [i], [ø] que [e] et [œ] que [ɛ]. L'arrondissement n'a par contre guère d'influence notable sur F_1. Enfin, traditionnellement F_3 est plus spécifiquement lié à la résonance de la cavité labiale. Ordonnées selon la fréquence croissante de F_3, on constate que [u]-[y] < [o]-[ø] < [ɔ]-[œ] < [a]-[ɛ] < [e] < [i]. [a] présente une valeur médiane (≈ 2500 Hz) entre la plus basse pour [u] (≈ 2000 Hz) et la plus haute pour [i] (≈ 3000 Hz). Au regard du geste labial, [a] peut être considérée comme une voyelle neutre, alors que [u] est produite avec la plus grande protusion et [i] avec le plus grand étirement, c'est-à-dire un écartement plus important des commissures des lèvres, assimilable à une rétraction labiale opposée au geste de protusion. Ainsi, F_3 baisse graduellement moins les voyelles sont étirées et plus elles sont arrondies, à savoir en fonction de la longueur croissante du résonateur labial. On peut également noter ici une interaction liant F_3 à F_1, ou plutôt à l'aperture des voyelles. En effet, s'agissant des voyelles arrondies [y ø œ u o ɔ], F_3 croît avec F_1. Inversement, pour les voyelles non arrondies [i e ɛ a], F_3 décroît avec F_1. Cette covariation inversée repose sur l'interférence de l'abaissement de la mandibule sur le geste labial. En effet, toute chose égale par ailleurs, plus la mâchoire inférieure est basse, plus l'amplitude de projection ou de rétraction des lèvres se réduit, du fait de la tension croissante qui s'exerce sur les tissus labiaux. Il est ainsi plus difficile d'arrondir ou d'étirer maximalement les lèvres plus la bouche est ouverte. Cela implique que plus les voyelles sont ouvertes, d'une part, moins elles sont étirées, et donc moins F_3 est élevé du fait d'un résonateur labial moins raccourci, et d'autre part, moins elles sont arrondies (projetées), moins F_3 est bas car le résonateur labial est moins allongé. La figure 1.5 illustre l'assez grande robustesse de ces paramètres formantiques face aux variations de style de parole (dialogue spontané *vs* lecture de mots isolés) et aux différences contextuelles de durée dues aux fluctuations locales de débit de parole ou à la variété de l'environnement prosodique ou segmental qui peuvent affecter les voyelles. Ainsi, on constate globalement que même si les valeurs brutes formantiques peuvent considérablement varier selon le contexte, relativement, les voyelles, les unes par rapport aux autres, se distinguent en moyenne assez bien selon les mêmes patrons F_1-F_2 protypiques exposés précédemment.

Avec le portugais, le français est la seule langue romane actuelle à compter des voyelles nasales lexicalement distinctives. Il présente par ailleurs l'autre particularité typologique de n'avoir aucunes voyelles nasales fermées, alors qu'elles composent

prioritairement (avec les ouvertes) les systèmes vocaliques où la nasalité est phonologique (Vallée 1994). Présentes dans des mots comme *Alain*, *alun*, *allons* et *allant*, les voyelles nasales du français, comptent trois mi-ouvertes : deux antérieures, [ɛ̃] non arrondie et [œ̃] arrondie, et une postérieure arrondie [ɔ̃], en sus de l'ouverte postérieure non arrondie [ɑ̃]. Ces voyelles sont assez souvent décrites dans les manuels de phonétique française (par exemple Marchal 1980, Tranel 1987) comme des répliques articulatoires nasalisées des voyelles orales correspondantes, respectivement [ɛ œ ɔ a]. La nasalité étant articulatoirement produite par un simple abaissement du voile du palais. Or, d'autres manuels (par exemple Carton 1974, Coveney 2001) et les travaux de Zerling (1984), Longchamp (1988) et plus récemment Delvaux (2002, 2003) montrent que ces voyelles présentent en fait des modifications assez systématiques dans leur configuration articulatoire linguale et labiale qui poussent à les considérer, non comme des voyelles orales simplement nasalisées, mais bien comme des unités vocaliques à part entière du système phonologique français. Ainsi, articulatoirement, face à leur pendant oral, les voyelles nasales sont globalement plus ouvertes, plus postérieures (plus variablement pour [ɔ̃]), et/ou plus arrondies (même [ɑ̃]). Ces différences articulatoires expliquent en partie la structure spectrale différenciée de ces voyelles nasales. Acoustiquement, les voyelles nasales apparaissent comme des sons éminemment complexes du fait de l'interaction importante de différents phénomènes. S'agissant tant de leur étude empirique en parole naturelle que de leur modélisation (Maeda 1993), ces voyelles présentent des complications acoustiques particulières qui rendent difficile une description précise et détaillée de leurs caractéristiques spectrales individuelles. Pour les appréhender assez globalement, il est nécessaire de distinguer les conséquences acoustiques de deux phénomènes : (i) l'abaissement du voile du palais, responsable de la nasalisation à proprement parler, interagissant avec (ii) des ajustements articulatoires spécifiques décrits ci-dessus. La nasalisation crée une bifurcation du tractus vocal à environ 10 cm de la glotte, mettant en communication les cavités nasale et orales. Cette configuration particulière du conduit entraîne une structure acoustique très complexe qui repose à la fois sur les résonances en parallèle de ces cavités, leur interaction et des antirésonances générées par l'embranchement dans le conduit. Ainsi, le spectre d'une voyelle nasale présente des éléments plus nombreux pour une même plage fréquentielle qu'une voyelle orale. Ces éléments constituent des paires de pôle-zéro acoustiques, appelés '(extra)formant' et 'antiformant' nasals, propres à la résonance de la cavité nasale et à son couplage aux cavités pharyngale et buccale. Ces paires d'extraformant-antiformant se superposent et affectent les formants issus de la résonance concomitante de la partie orale du conduit vocal. Le modèle à tubes permet théoriquement de calculer, par la somme des fonctions de transfert couplées des conduits oral et nasal, la fonction de transfert de ces configurations géométriques complexes associées aux différentes voyelles nasales. Reste que ces simulations se heurtent à une variabilité importante observées empiriquement et encore difficilement prédictible à ce jour. Cette indétermination tient, en premier lieu, aux différences morphologiques importantes des fosses nasales et des sinus (dont le rôle reste obscur) très variables d'un individu à l'autre. Ainsi, si certaines paires de formant-antiformant caractéristiques de la fonction de transfert de la cavité nasale d'un individu donné sont stables quelle que soit la voyelle réalisée, elles diffèrent largement inter-individuellement. En second lieu, la sortie acoustique est très sensible au niveau de couplage naso-pharyngal et naso-buccal

assujettis au degré d'abaissement du voile. Or, l'ouverture vélo-pharyngée se montre peu constante d'un individu à l'autre, d'une production à l'autre et d'une voyelle à l'autre, même si certaines régularités sont attestées, notamment entre aperture vocalique et abaissement vertical du voile (Delvaux 2003, Amelot 2004). On sait que la qualité de ce couplage est responsable d'une bonne partie des propriétés spectrales fondamentales des voyelles nasales. Par exemple, Maeda (1993) montre que plus ce couplage est important, plus formants et antiformants d'une même paire divergent. Ces indéterminations rendent donc assez aléatoires les prédictions relatives aux interférences entre les formants oraux et les formants-antiformants nasals, pouvant produire des effets variés et combinables : élargissement de la bande d'un formant, atténuation de son énergie, fusion formantique, ou les trois à la fois. Aussi, la détection et les mesures spectrales des voyelles nasales naturelles montrent des difficultés spécifiques. Malgré ces obstacles, des régularités robustes se dégagent des analyses acoustiques sur les voyelles nasales du français. L'étude de Delvaux (2003) est particulièrement informative du fait qu'elle isole les effets du couplage naso-oral, relatif au geste d'abaissement du palais mou responsable de nasalisation des voyelles, des effets induits par les configurations labio-linguales distinctes des voyelles orales correspondantes. Acoustiquement, la nasalisation montre des conséquences systématiques pouvant affecter tout le spectre des voyelles, mais plus particulièrement concentrées dans les zones de F_1 et F_3. Le premier effet concerne une chute conséquente de l'énergie globale croissante avec la fréquence, explicable par l'activité d'antirésonance et par une impédance plus importante du conduit du fait d'un volume d'air et d'une surface de résonance plus importants dispersant l'énergie vibratoire. Plus particulièrement, cette perte se concentre sur une bande spectrale inférieure à 1000 Hz, à savoir celle principalement concernée par F_1, et secondairement par F_2 pour les voyelles postérieures. Parallèlement, la bande fréquentielle 2000-3000 Hz, intéressée par F_3, est aussi le lieu d'une forte atténuation de l'énergie du fait probable d'antirésonances importantes. Acoustiquement, la nasalité des voyelles se traduit donc tout d'abord par la baisse d'énergie décrite pour la nasalisation. En lien avec les différences articulatoires constatées par rapport à leur pendant oral, deux autres caractéristiques de la nasalité vocalique en français sont rapportées. D'une part, une élévation du F_1 est constatée en relation avec une articulation plus ouverte des voyelles nasales. D'autre part, un abaissement du F_2 des voyelles antérieures non arrondies [ɛ̃ ɑ̃] est observé en conséquence de leur caractère plus postérieur et/ou plus arrondi. Les différences avec les voyelles orales correspondantes reposeraient ainsi en bonne partie sur un repositionnement articulatoire de la langue et/ou des lèvres. Enfin, il est noté que les voyelles nasales antérieures se particularisent également par la présence d'une proéminence plus intense de F_2, autour de 1300 Hz, et donc hors des zones d'atténuation provoquées par la nasalisation. A titre d'exemple, la figure 1.6 reproduit les différences fréquentielles de F_1, F_2 et F_3 observées par Delvaux (2002) entre voyelles nasales et orales sur deux locuteurs belges francophones. Les investigations perceptives complémentaires réalisées par Delvaux (2003) montrent par ailleurs que la conjugaison de la baisse d'énergie spectrale et l'abaissement de F_2 sont essentiels pour satisfaire l'identification des voyelles nasales en français.

1.3.2 *Les consonnes du français*

Articulatoirement, la parole apparaît globalement comme une alternance régulière d'ouvertures plus ou moins grandes et de fermetures plus ou moins complètes du conduit vocal. Ce cycle d'oscillations ouverture-fermeture est principalement supporté par l'abaissement et l'élévation rythmés de la mâchoire inférieure. MacNeilage (1998) pose cette modulation cyclique mandibulaire comme le substrat physiologique de la structure organisationnelle primitive de la parole : la syllabe[8]. La phase d'ouverture est généralement assurée par les articulations vocaliques et la phase de fermeture par les productions consonantiques. Acoustiquement, ce cycle articulatoire a pour conséquence une fluctuation régulière de l'intensité du signal de parole, plus importante dans les phases d'ouverture, correspondant aux voyelles, que dans les phases de fermeture du conduit, relatives aux consonnes. Ainsi, toutes choses égale par ailleurs, plus le conduit vocal est ouvert plus le son est fort ; plus il est fermé, moins l'énergie acoustique irradie hors du conduit, plus l'impédance acoustique du conduit est grande, et donc moins le son en sortie est intense. Par exemple, si l'on demande de crier le plus fort possible, il y a de grandes chances que le son [a] soit émis instinctivement, plutôt que [u], et certainement pas [m]. La raison en est que [a] est le son réalisé avec le conduit le plus largement ouvert, beaucoup plus que pour [u], alors que [m] est produit bouche fermée. Les consonnes sont donc des sons qui à la base se distinguent acoustiquement des voyelles par une énergie sonore moins importante, concomitante à leur degré de fermeture plus important du conduit vocal. Cette perte d'énergie acoustique lors de l'émission de consonnes est clairement visible sur un oscillogramme ou un spectrogramme (figures 1.7 et 1.8) où comparée aux voyelles [a] précédente et suivante la consonne montre une amplitude acoustique significativement réduite : l'onde est verticalement moins ample sur l'oscillogramme et le tracé fréquentiel moins sombre sur le spectrogramme. A titre d'exemple, Tubach (1989) note une perte de 3 à 35 dB pour les fricatives du français. Cette discontinuité d'amplitude constitue un indice spectral majeur de rupture dans le signal de parole utile pour segmenter les consonnes.

En sus, la fermeture du conduit a également pour conséquence d'atténuer l'énergie propre de la source glottique, à savoir l'intensité du voisement. Le voisement est visible sur un spectrogramme par une barre horizontale plus ou moins sombre à hauteur de la F0 (généralement ≤ 300 Hz) et parfois peu distinguable de F_1, voire fusionné avec dans le cas des voyelles fermées caractérisées par un F_1 très bas. Cet indice spectral apparaît clairement par exemple pour [b] (voisé) comparé à [p] (non voisé) dans la figure 1.8. Outre les consonnes sourdes pour lesquelles l'énergie glottique est nulle du fait de l'absence de vibrations des cordes vocales, les consonnes voisées montrent toutes une baisse de l'énergie du ton laryngé avec des degrés divers selon le type de consonnes. Plus les consonnes sont fermées, plus la perte d'énergie glottique est importante. Cela tient directement aux conséquences de la fermeture supraglottique réalisée dans le conduit vocal sur l'amplitude des vibrations des cordes vocales. A l'inverse des

[8] Pour une revue de la notion de syllabe en phonétique et en phonologie, voir notamment Meynadier (2001), Rousset (2004) et Ridouane et al. (2011).

consonnes sonantes [j ɥ w l m n ɲ ŋ] pour lesquelles le conduit oral ou nasal reste assez largement ouvert, les obstruantes voisées [b d g v z ʒ ʁ] (comme les sourdes [p t k f s ʃ]) sont articulatoirement réalisées soit par une fermeture complète du conduit vocal dans les cas des plosives [b d ɡ], soit par une constriction très étroite dans le cas des fricatives [v z ʒ ʁ]. Cette fermeture drastique provoque une accumulation de l'air expiré en amont de la constriction, d'où une augmentation de la pression aérodynamique à l'intérieur du conduit vocal, appelée 'pression intra-orale' (PIO). Or, l'établissement, le maintien et l'amplitude de la vibration des cordes vocales sont conditionnés par la puissance du flux d'air expiratoire forçant le passage des cordes vocales accolées. Un flux d'air se crée d'une zone de pression plus élevée vers une zone de pression moins élevée. Pour qu'il y ait voisement, il faut donc que la pression sous-glottique (PSG) de l'air venant de la trachée soit toujours supérieure à la PIO. Cette différence, appelée 'différentiel transglottique de pression', doit au minimum se situer autour de 2 hPa, sinon les cordes vocales ne peuvent être mises en vibration par le flux d'air expiratoire. En outre, plus ce différentiel est important, plus le flux d'air transglottique est puissant, donc plus l'amplitude de vibration est grande, et par là-même plus l'énergie acoustique résultante intense. Dans le cas des consonnes sonantes, l'air pouvant s'échapper plus librement du fait d'un conduit plus ouvert, une PIO nulle ou quasi nulle assure un différentiel toujours important, ce qui constitue une condition de voisement optimale équivalente à celle observée pour les voyelles. Dans le cas des consonnes obstruantes, l'air expiré venant de la glotte étant totalement ou partiellement retenu par la fermeture supraglottique du tractus vocal, la PIO augmente rapidement dans le conduit. La PSG étant elle relativement constante au cours d'un énoncé (et même décroissante au cours d'un énoncé long), le différentiel transglottique de pression diminue d'autant que la PIO croît au cours de la consonne. Cela provoque alors mécaniquement une réduction de l'amplitude de vibration des cordes vocales, et par là-même de l'intensité de l'onde glottique à la source du voisement de la consonne. Aussi, la barre de voisement des consonnes obstruantes voisées (figure 1.8) apparaît distinctement plus faible que celles des consonnes sonantes (figure 1.7).

Plus globalement, sonantes et obstruantes se distinguent par une répartition différente de l'énergie dans leur spectre. Il est ainsi noté que l'énergie spectrale des obstruantes se distribue essentiellement au-delà de 2-3000 Hz, d'autant plus que la consonne est sourde. En-deçà, l'énergie acoustique est quasi nulle, ou faible dans le cas d'obstruantes voisées. Cette énergie spécifique des fréquences plus hautes est le fait de la source supraglottique générant le bruit constitutif de la consonne obstruante, à lui seul ou en superposition à l'émission glottique voisée. Comme vu précédemment, ce bruit provient de l'échappement turbulent du flux d'air expiratoire brusquement relâché après une occlusion (plosives) ou très perturbé par une constriction critique (fricatives), réalisée en différents points du conduit vocal : lèvres, dents/alvéoles, palais dur ou mou, luette… Il est communément admis (Fant 1960, Harrington & Cassidy 1999 notamment) que la cavité en aval de la constriction participe pour l'essentiel à la résonance de ces bruits supraglottiques. Cela peut aisément s'observer en produisant par exemple [sssss] en arrondissant et désarrondissant alternativement les lèvres. La participation de la cavité postérieure au point constriction, décrite comme plus mineure, occasionnerait des résonnances sous 4000 Hz (Harrington & Cassidy 1999). Même si sa contribution à la

forme acoustique des obstruantes reste assez largement à préciser[9], on peut empiriquement la constater par exemple en produisant isolément un [p] tout en positionnant la langue comme pour réaliser un [y] *vs* un [u] subséquent. L'explosion de [p] est perceptiblement plus grave quand la cavité buccale est plus longue comme pour la voyelle postérieure [u], que quand elle est plus courte comme pour la voyelle antérieure [y]. Cette asymétrie en faveur des résonances issues de la cavité avant dans le cas des obstruantes est intuitivement compréhensible par le fait que produire une constriction très importante dans le conduit a pour conséquence d'augmenter la vélocité du flux d'air par l'augmentation de la PIO. Constriction très étroite plus flux d'air sortant plus véloce réduisent drastiquement les possibilités d'excitation d'un résonateur situé en direction inverse au flux d'air et en arrière de la source acoustique, participant par là-même à découpler en bonne partie la cavité arrière de la source bruitée. En français[10], bruits d'explosion ou de friction, plus ou moins intenses selon les consonnes, sont toujours générés au-dessus de la glotte, inversement au ton laryngé produit à ce niveau de l'appareil vocal. Dès lors, la cavité avant, étant toujours plus petite lors de bruits supraglottiques, favorise la résonance de fréquences plus hautes. De faibles fréquences basses émergent essentiellement dans le cas des obstruantes voisées. Par contre, les sonantes, phonologiquement voisées et non bruitées, montrent une énergie plus importante sous 3-4000 Hz et faiblissant graduellement dans les fréquences élevées. Elles sont en cela plus proches des voyelles, dont elles partagent d'autres caractéristiques, notamment leur structure formantique basée sur la résonance du ton périodique laryngé dans tout le conduit vocal. Ainsi, la distribution spectrale de l'énergie associée au type de source(s) acoustique(s) permettent de distinguer la catégorie majeure des consonnes obstruantes, et entre elles les voisées des non voisées, de celle des sonantes, par défaut voisées.

Les consonnes sonantes du français se subdivisent en deux catégories principales : les approximantes et les nasales (figure 1.7). Nous ne nous attardons pas ici sur une troisième catégorie de consonnes sonantes : les vibrantes (ou 'roulées') alvéolaire et vélaire ne constituant en français que des variantes régionales assez locales ou idiolectales de /ʁ/. Elles sont réalisées par la mise en vibration soit de l'apex de la langue au niveau des alvéoles (petits bourrelets situés derrière les incisives supérieures) pour [r], soit de la luette (ou uvula) sur le dos de la langue pour [ʀ][11]. Leur image spectrographique (figure 1.7) rend compte de cette articulation par la présence de stries verticales claires au cours de la consonne correspondant à la chute d'énergie due aux micro-occlusions issues du battement répété de l'articulateur. Les consonnes approximantes [j ɥ w l] présentent les caractéristiques acoustiques les plus proches des voyelles, à savoir une structure formantique riche. Elles posent dès lors de réelles difficultés de segmentation dans la parole continue notamment en contexte

[9] Voir notamment sur ce point Badin (1991) et les travaux récents de Toda (2009).

[10] En dehors du coup glotte [ʔ], parfois produit lors d'accent d'insistance ou afin d'éviter certains enchaînements (hiatus, liaison sans enchaînement, désambigüisation…).

[11] On trouve aussi cette réalisation moins rarement après une obstruante voisée (Meunier 1994), d'autant plus que la syllabe porte un accent d'insistance.

intervocalique. Le meilleur indice visuel est la baisse d'intensité particulièrement dans les zones de F_2 à F_4 (500-4000 Hz), lié à la fermeture un peu plus importante du conduit face aux voyelles. Les approximantes centrales [j ɥ w] sont ainsi communément appelées 'semi-consonnes' ou 'semi-voyelles' du fait de leur grande similarité articulatoire et acoustique avec les voyelles correspondantes [i y u], présentant ainsi un F_1, F_2 et F_3 aux mêmes fréquences que ces dernières. Elles s'en distinguent par contre par une durée plus courte et par une instabilité des trajectoires formantiques pouvant ne présenter aucun état stable, ce qui rend souvent difficile la détermination de leur limite temporelle. C'est ce caractère dynamique de leur structure spectrale qui les différencie donc aussi des voyelles hautes. Il tient à une articulation réalisée par des gestes – de la lame de la langue vers le palais dur pour la palatale [j], combiné à l'arrondissement des lèvres pour la labio-palatale [ɥ] et du dos de la langue vers le voile associé à l'arrondissement pour la labio-vélaire [w] – qualifiés de 'balistiques' ou de 'transitoires' ne montrant bien souvent aucune phase de tenue de la constriction : dès la cible atteinte ou approchée l'articulateur s'en éloigne aussitôt. La sonante alvéolaire [l] est la seule approximante latérale du français. Elle est articulée par une occlusion incomplète effectuée par l'apex sur les alvéoles combinée à un abaissement important des côtés de langue permettant un échappement latéral et silencieux de l'air phonatoire. La structure formantique de cette consonne est plus complexe du fait de présence d'antiformants issus de la bifurcation latérale du conduit vocal provoqué par l'obstacle apical. Ces antirésonances apparaissent principalement au niveau de F_4, mais peuvent également annihiler complètement F_3 et F_2 en fonction du contexte vocalique. Ce dernier exerce aussi une forte influence sur les valeurs-mêmes de F_2, variant de 1900 à 1300 Hz selon que les voyelles contigües sont antérieures ou postérieures respectivement, et de F_3, flottant entre 2700 Hz (en contexte [i]) et 1400 Hz (en contexte [u]) selon que les voyelles environnantes sont arrondies ou non. F_1 est par contre beaucoup plus stable autour de 300 Hz (Tubach 1989), ce qui, par le décrochage vers le bas de F_1, procure un bon indice de segmentation pour cette consonne en contexte vocalique non fermé.

Les consonnes sonantes nasales [m n ɲ ŋ] se caractérisent également par une structure formantique très complexe impliquant des formants oraux et nasals et des antiformants, issus de la division du conduit vocal en sortie de la cavité pharyngale entre voie nasale et voie orale, provoquée par l'abaissement du voile du palais. Hormis la position basse du velum, l'articulation des nasales repose sur l'occlusion du conduit oral effectuée par les lèvres pour [m], par l'apex et/ou la lame de langue sur la face interne des incises supérieures et/ou les alvéoles pour [n], par le dos de la langue au niveau du palais dur pour [ɲ] ou sur le palais mou pour [ŋ]. Leur configuration articulatoire orale est similaire à celle des consonnes plosives de même lieu : bilabial, dento-alvéolaire et vélaire (le français ne comptant pas de plosive palatale). [ɲ ŋ] occupent une place particulière dans le système du français du fait de leur distribution plus contrainte, de leur occurrence lexicale particulière et de leur articulation actuelle. [ɲ] n'est attesté (par exemple dans le dictionnaire Le Robert) en initiale de mot que dans six de mots assez peu fréquents d'origine régionale, argotique ou étrangère : « gnon, gnôle, gniouf, gnangnan, gnognotte, gnocchi ». De plus son articulation contemporaine est largement décrite comme celle d'une séquence consonantique [nj] impliquant une occlusion

apicale [n] (probablement palatalisée [nʲ]) suivie à son relâchement d'une articulation approximante dorso-palatale [j] (souvent en tout ou partie nasalisé [j̃]), et non plus comme la véritable consonne palatale pleine [ɲ] (Bothorel et al. 1986, Coveney 2001). Outre en méridional où il est fréquemment réalisé en appendice de voyelles nasales finales de mot et dans le cas d'assimilation nasale de [g] entre voyelle nasale et consonne, [ŋ] apparaît en français exclusivement en fin de mots empruntés récemment à l'anglais. Il montre également une désarticulation en une séquence [ŋg] ou en une nasale vélaire partiellement dénasalisée [ŋᵍ] (Coveney 2001). Du point de vue acoustique, le modèle de Fant (1960) prédit pour toute consonne nasale un premier formant nasal (N_1) stable vers 250-400 Hz et un intervalle régulier de 700 à 800 Hz entre les formants nasals supérieurs (d'où $N_2 \approx$ 950-1200 Hz, $N_3 \approx$ 1650-2000 Hz, etc.). Ces formants, quoique variables selon les morphologies individuelles, sont identiques quel que soit le lieu d'articulation de la nasale, car issus de la résonance du tube naso-pharyngal dont la taille est très constante pour un sujet donné. Par contre, la cavité buccale, fonctionnant en parallèle comme un résonateur annexe fermé, introduit des antirésonances dans le spectre du son qui sont plus variables et dépendants du lieu d'articulation de la consonne. S'il n'existe pas réellement de moyen fiable d'estimer la fréquence précise de ces antiformants, on sait quand même qu'ils répondent pour l'essentiel à la longueur du résonateur oral fermé. Plus celui-ci est long, plus bas sont les antiformants. Ainsi, [m] montre un antiformant vers 1400 Hz, alors que celui de [n] se situe plutôt vers 1800 Hz : son lieu d'articulation plus arrière réduit d'autant la longueur du tube oral, ce qui favorise un zéro spectral plus haut (Harrington & Cassidy 1999, Tubach 1989). Enfin, le spectre des consonnes nasales se composent aussi de formants propres à la résonance du conduit oral. Les formants oraux sont généralement dans le prolongement des formants vocaliques, mais sont souvent très faibles du fait de la proximité fréquentielle des antiformants. Comme [l], les consonnes nasales présentent un F_1 stable vers 300 Hz qui fournit là aussi un indice de segmentation utile en contigüité aux voyelles non fermées. La distinction perceptive des lieux d'articulation entre les différentes consonnes nasales reposerait essentiellement sur la trajectoire suivie par les formants (principalement F_2 et F_3) dans les enchaînements avec la voyelle précédente et/ou suivante, appelée 'transition formantique' (cf. ci-après).

Inversement aux sonantes qui se caractérisent par une structure formantique et périodique, les consonnes obstruantes s'en distinguent clairement par la présence de bruit (source apériodique généré dans le conduit vocal) et de faibles composantes formantiques. On compte deux catégories d'obstruantes en français[12] : les plosives et les fricatives (figure 1.8). Étant non nasales, le voile relevé exclut toute participation du conduit nasal à leur forme spectrale. Néanmoins, elles présentent, elles aussi, des antirésonances du fait de l'existence d'une source acoustique supraglottique. Liées à l'activité du résonateur arrière, celles-ci concerneraient plus spécifiquement le bas du spectre. Ainsi, la chute de l'énergie constatée pour les fricatives [s] ou [ʃ] à une fréquence de coupure située respectivement vers 4000 Hz et 2000 Hz en est caractéristique (Harrington & Cassidy 1999, Toda 2009). Reste que ce phénomène

[12] Nous laisserons ici de côté les consonnes affriquées au statut discuté en français.

semble n'apporter qu'une contribution plus mineure à la caractérisation du patron spectral des obstruantes pour lesquelles le résonateur avant joue le premier rôle dans l'émergence des pics spectraux essentiels à la reconnaissance de ces consonnes.

Les fricatives sont articulées avec une constriction critique du conduit vocal : entre la lèvre inférieure et les incisives supérieures pour les labio-dentales [f v], entre la pointe de la langue et les alvéoles pour les alvéolaires [s z], entre la lame et l'arrière des alvéoles pour les postalvéolaires (ou alvéo-palatales) [ʃ ʒ][13], et entre le dos et la luette pour l'uvulaire [ʁ]. Ce rétrécissement mettant en turbulence l'air expiré, produit un bruit dont l'amplitude et la structure fréquentielle dépendent de la pression de l'air dans l'aire de constriction, de son diamètre et de la localisation d'obstacles à la colonne d'air en sortie du chenal fricatif (Fant 1960). Ces facteurs permettent de séparer les fricatives dites 'sibilantes' des autres. Les sibilantes montrent une intensité de bruit de friction plus importante, car le jet d'air expulsé percute l'obstacle à angle droit proposé par les incisives supérieures pour [s z] et [ʃ ʒ], ce qui a pour effet d'en augmenter la turbulence (Shadle 1990). Par contre, l'obstacle bien moins frontal au flux d'air de la lèvre inférieure pour [f v] ou du palais pour [ʁ] n'amplifie guère le bruit. Ainsi, [s] et [ʃ] ont une intensité globale supérieure de 10 à 30 dB à [f] et à la variante sourde [χ] de /ʁ/ (Tubach 1989). La répartition spectrale de l'énergie du bruit est également différente en fonction du lieu d'articulation déterminant la taille du résonateur en aval de la source de bruit. Les sibilantes présentent une énergie plus polarisée sur une bande fréquentielle spécifique que les non sibilantes montrant une distribution plus diffuse. Ainsi, un pôle d'énergie maximale vers 2000-4000 Hz émerge typiquement pour [ʃ ʒ], alors que celui de [s z], articulés plus en avant, se situe vers 4000-7000 Hz. Pour [f v], du fait de l'antériorité plus extrême de leur articulation, on attendrait un bruit fricatif polarisé dans des fréquences encore plus hautes. Or, l'absence de véritable obstacle et de cavité de résonance en avant de la constriction conditionne un bruit faible et sans relief coupé dans les basses fréquences. Comme déjà évoqué, les fricatives voisées diffèrent des sourdes par une énergie globale plus importante dans les basses fréquences. Par exemple, Tubach (1989) rapporte une diminution par rapport aux voyelles de 20 à 30 dB pour [v] et de 5 à 10 dB pour [ʁ] comparée à une perte de 25 à 35 dB pour [f] et de 15 à 20 dB pour [χ]. L'énergie supérieure des voisées provient de leur source glottique périodique qui occasionne notamment l'apparition de formants plus ou moins faibles, principalement dans le prolongement du F_2 et F_3 des voyelles contiguës. Par contre, les voisées montrent systématiquement un bruit fricatif moins intense, aisément observable dans la figure 1.8. L'atténuation du bruit fricatif est mécaniquement due à la résistance posée par la fermeture de la glotte nécessaire à la vibration des cordes vocales. Cet obstacle réduit drastiquement le flux d'air expiratoire passant dans le conduit supraglottique, ce qui affaiblit d'autant la valeur de la PIO. La PIO bien plus faible des voisées engendre des turbulences moins importantes et donc un bruit en sortie de la constriction supraglottique bien moins intense. Reste la fricative uvulaire /ʁ/ du français dont il est difficile de proposer une image simple et homogène. Outre les variantes libres

[13] Voir Toda (2009) pour une étude détaillée des différentes stratégies articulatoires dans la production de ces sons en français.

sonantes vibrantes [r ʀ] décrites précédemment, ce phonème est très polymorphe en français (Meunier 1994) du fait d'une grande sensibilité au contexte vocalique et consonantique. Par défaut fricatif uvulaire voisé, /ʁ/ peut être aussi : (i) après obstruantes sourdes ou en fin de mot, totalement sourd [χ] ou partiellement dévoisé [ʁ̥], (ii) entre deux voyelles, plus ouvert avec très peu ou sans bruit fricatif telle une approximante [ʁ̞], (iii) sous un accent d'insistance, plus fermé avec superposition de bruit et de battements, voisé [ʀ] après obstruantes voisées ou dévoisé [ʀ̥] après obstruantes sourdes. Cette grande variabilité et son intensité parfois très faible, notamment en fin de mot et en intervocalique, en rendent souvent la segmentation acoustique bien délicate. S'agissant des réalisations nettement fricatives, la fréquence du bruit de [ʁ], et de sa variante non voisée [χ], peut être différente selon que son lieu d'articulation avance en zone vélaire ou pré-vélaire (proche de [x ɣ]) au contact de voyelles antérieures, ou qu'il se maintienne en zone uvulaire en contexte de voyelles postérieures. Reste que globalement [χ] montre un bruit généralement diffus et peu intense pouvant laisser apparaître de faibles pôles dans les fréquences moyennes ou basses selon la nature du contexte vocalique, du fait de la résonance d'une cavité avant très volumineuse. Dans le cas de [ʁ], la résonance additionnelle du ton glottique favorise l'émergence d'une structure formantique dans le bas du spectre et un affaiblissement important ou un masquage parfois quasi total du bruit de friction.

Les consonnes plosives présentent la rupture la plus franche avec les voyelles. Elles sont réalisées par la fermeture totale des conduits nasal et oral au lieu bilabial pour [p b], dento-alvéolaire pour [t d] et vélaire pour [k g]. La tenue de cette occlusion se traduit acoustiquement par un silence (absence totale d'énergie sonore) durant la majeure partie des plosives sourdes. Pendant cette même phase, les voisées présentent une barre de voisement d'intensité faible et décroissante du fait de l'augmentation de la PIO au cours de consonne. La résonance étouffée du ton laryngé dans le conduit fermé peut néanmoins parfois fournir une très faible structure formantique à proximité des formats vocaliques (visible par exemple pour [g], figure 1.8). Le relâchement brusque (quelques ms) de l'occlusion se traduit acoustiquement par une explosion correspondant à une fine barre foncée verticale sur un spectrogramme. Cette explosion est généralement suivie par un court bruit de friction le temps (5 à 35 ms) que l'articulateur s'écarte suffisamment des parois du conduit (lèvres, dents, alvéoles ou voile) pour ne plus produire une constriction critique à l'origine de ce bruit supraglottique. On peut ainsi aisément observer sur la figure 1.8 que cette phase est plus longue pour les plosives dorsales [k g], qu'apicales [t d] et que bilabiales [p b], tenant au fait que plus l'articulateur est véloce (lèvres > apex > dos), plus le bruit d'explosion et de friction est bref. Comme pour les fricatives, bruits d'explosion et de friction sont plus faibles pour les voisés du fait d'une PIO moins importante consécutive à la fermeture glottique. Leur faible explosion étant parfois invisible sur un spectrogramme, [b d g] se distinguent alors globalement des nasales ou de [l] par une intensité générale plus faible. Alors que le silence ou le voisement lors de la tenue n'apporte aucun indice sur le lieu d'articulation, la structure fréquentielle du bruit émis lors du relâchement de la plosive en dépend. Il répond aux mêmes caractéristiques de résonance que celui des fricatives de même lieu, à la différence près qu'il est non maintenu et donc beaucoup plus bref. Ainsi, le spectre du

bruit des bilabiales est distribué sur une large plage fréquentielle, d'intensité faible et décroissante plus les fréquences sont élevées. Les plosives dento-alvéolaires se caractérisent par un bruit d'intensité plus importante et croissante avec les fréquences. Présentant généralement une énergie maximale entre 3000 et 6000 Hz, il est plus aigu. Enfin, les plosives vélaires, comme la fricative uvulaire, sont difficiles à caractériser, car très influencées par l'environnement vocalique. Pré-vélaires en contact de voyelles antérieures, leur spectre est proche de celui d'alvéolaires : croissant (dit 'montant') et aigu ; post-vélaires en contexte vocalique postérieur, leur spectre tend à ressembler à celui des bilabiales : sans véritable pôle (dit 'plat') ou décroissant (dit 'descendant') et grave. Typiquement, il est décrit avec deux pôles spectraux variables entre 1500 et 4000 Hz, suivi d'un bruit de friction intense et long du fait de l'inertie plus importante du dos de la langue. Un pincement entre le F_2 et le F_3 des voyelles avant/après constitue généralement un indice spectrographique assez net de leur lieu d'articulation vélaire (Harington & Cassidy 1999), comme cela est clairement illustré par les figures 1.8. et 1.9a.

Ce dernier point est crucial pour l'identification acoustique et perceptive du lieu d'articulation des consonnes. En effet, toutes les informations de lieu ne sont pas internes à la consonne. Les indices acoustiques les plus saillants sont portés par les sons voisins. Principalement, les formants des voyelles enchaînées aux consonnes dessinent des pentes caractéristiques (figure 1.9a). Depuis les travaux fondateurs du Haskins Laboratory dans les années 50 (notamment Liberman et al. 1954), on sait que les transitions formantiques vocaliques sont essentielles à la perception du lieu d'articulation de la consonne, qu'elle soit sonante ou obstruante. Faisant écouter des séquences [apˀ], [atˀ] ou [akˀ] tronquées avant le bruit d'explosion, l'auditeur n'a généralement aucun mal à reconnaître la consonne, car principalement les F_2 et F_3 de la voyelle ont une direction et/ou degré de pente différents aux abords de chacun de ces consonnes[14]. Ainsi, du relâchement de la consonne à la valeur formantique de la partie stable de la voyelle (cible vocalique), les transitions sont selon les voyelles (i) dans le cas des bilabiales plates ou montantes pour F_2 et F_3 ; (ii) dans le cas des alvéolaires montante devant des voyelles antérieures et descendante avant des voyelles postérieures ou [a] pour F_2 et descendante pour F_3 ; (iii) dans le cas des vélaires plate ou descendante pour F_2 et plate ou montante pour F_3 (Tubach 1989). Que la consonne soit sonante, orale ou nasale, ou obstruante, plosive ou fricative, les transitions sont similaires pour chaque lieu d'articulation. De ces observations, Delattre et al. (1955) en ont formulé la *théorie du locus*, à savoir d'une cible acoustique consonantique vers laquelle convergent les formants des voyelles contigües. Le locus de F_2 est principalement impliqué dans l'identification du lieu d'articulation consonantique. Il se situerait vers 700 Hz pour les labiales et 1800 Hz pour les alvéolaires. Les vélaires présenteraient deux locus : un vers 3000 Hz pour les voyelles antérieures et l'autre vers 1000 Hz pour les postérieures (figure 1.9b). Les mesures acoustiques en parole continue ne confirment qu'en partie cette théorie. Ainsi, Harington & Cassidy (1999) retirent de travaux ultérieurs que si le locus alvéolaire est généralement assez robuste, les labiales en présenteraient au moins

[14] Les travaux de Strange (1989) ont montré que ces transitions sont également essentielles à la reconnaissance des voyelles, tout autant ou plus que leur partie stable.

deux différents selon les voyelles et ceux des vélaires restent très difficiles à mettre en évidence au regard de la grande sensibilité du lieu de ces consonnes face au contexte vocalique. Reste que les transitions formantiques rendent compte d'une continuité dans les processus de production de la chaîne sonore. Les unités de la parole n'y sont pas juxtaposées l'une après l'autre et ne montrent pas de frontière ou de rupture très nette. Au contraire la parole continue repose sur des interpolations successives entre différents domaines spectraux enchaînés, parfaitement illustrées par les tracés spectrographiques où les transitions entre sons se font de façon plus continue qu'abrupte. Ces interpolations sont les conséquences acoustiques du processus de coproduction articulatoire intrinsèque à la production de la parole continue.

1.3.3 *Coarticulation et réduction en français*

La langue se caractérise par des processus fondamentalement dynamiques qui président à la structuration des systèmes d'unités sonores distinctives (phonèmes)[15] et à leur réalisation en unités phonétiques contrastives dans la parole (sons). Le système sonore d'une langue répond à une variation tant diachronique que synchronique : régionale, stylistique ou individuelle. Chez un locuteur donné, un même phonème montre des formes sonores différentes en parole continue selon sa position dans la syllabe, le mot ou le syntagme, son caractère accentué ou non, et surtout son environnement phonétique lié à la nature des sons voisins enchaînés. Ce dernier facteur de variation de prononciation des unités phonémiques met en jeu des mécanismes de coproduction des propriétés articulatoires entre sons contigus dans la chaîne parlée. En effet dans l'enchaînement de sons, certains gestes articulatoires sont anticipés et/ou peuvent persévérer lors de la réalisation des phonèmes précédents et/ou subséquents. Ceux-ci portent alors les traces acoustiques de ces chevauchements articulatoires. Référés sous le terme de 'coarticulation', ces indices jalonnent le signal acoustique. Ainsi, les transitions formantiques des voyelles en sont les marques les plus apparentes, rendant compte du passage d'une position articulatoire à une autre de manière continue et non discrète.

Les processus de coproduction sont intrinsèques et nécessaires à l'acte de parole. Le chevauchement des articulations optimise en effet la transmission des messages linguistiques en permettant l'émission d'informations phonétiques de plusieurs sons en parallèle, puisque divers gestes peuvent être produits simultanément (Mattingly 1981). La coarticulation accroît la vitesse de transmission, plusieurs de sons chevauchés étant réalisés dans un laps de temps plus court que s'ils étaient produits de manière strictement séquentielle. De manière équivalente, on produit plus de battements en un

[15] Nombre de travaux et de théories de la parole démontrent la place centrale occupée par des principes dynamiques acoustico-perceptifs et articulatoires et leurs interactions dans l'émergence et l'architecture des systèmes phonologiques. Ces questions sortant du cadre de ce chapitre, nous invitons le lecteur à se reporter, entre autres, à la *théorie quantique* de Stevens (1989), la théorie de la *dispersion adaptative* de Lindblom (1986), ou la théorie de la *perception pour le contrôle de l'action* de Schwartz et al. (à paraître).

temps donné, si on alterne entre deux doigts, que si l'on les réalise avec un seul : le mouvement d'un doigt pouvant être initié sans attendre la fin de l'autre. Conjointement, la coarticulation est le produit de processus physiologiques reposant en partie sur l'inertie biomécanique des articulateurs et la force, ou énergie, déployée pour leur déplacement. Ces mécanismes biologiques, comme pour tout autre comportement du vivant, sont en bonne partie guidés par un principe d'économie énergétique qui sous-tend des organisations articulatoires adaptées aux demandes de la communication : satisfaire un résultat acceptable en limitant les moyens et les efforts mobilisés. Ainsi, face à une nécessité d'augmenter son débit d'élocution (par exemple, si on sait que la conversation peut s'arrêter brusquement), un locuteur à deux stratégies possibles en fonction d'enjeux communicationnels (exprimables en termes d'intelligibilité). Soit il augmente d'autant la vitesse de déplacement de ces articulateurs afin de tenter d'atteindre malgré tout les cibles linguistiques, maximisant ainsi leurs caractéristiques discriminantes et leur intelligibilité 'individuelle'. Cette précision articulatoire a cependant un coût plus élevé en termes d'effort fourni. Soit s'il 'veut' plutôt limiter son effort articulatoire, il peut alors réduire l'amplitude et/ou la vélocité de ses mouvements articulatoires, ce qui ne permet que d'approcher les cibles linguistiques. Les sons produits sont alors, par perte ou amenuisement de certaines de leurs propriétés caractéristiques, paradigmatiquement moins distinctifs et/ou syntagmatiquement moins contrastifs avec les sons voisins dans la chaîne parlée, mais suffisamment pour préserver une compréhensibilité contextuelle et/ou plus globale de la parole. Dans ce cas, on peut aussi observer une réorganisation articulatoire pouvant impliquer des stratégies de compensation, visant à pallier la perte d'information par la réalisation d'un geste différent plus économique et satisfait une équivalence acoustico-perceptive de la propriété détériorée (Kohler 1992, 1995). La coarticulation apparaît donc comme un moyen adaptatif d'optimisation complexe de production de la parole sur lequel le locuteur exerce un contrôle (économie d'effort, réorganisation spatiale et temporelle des gestes…) en fonction d'objectifs communicationnels liés au degré d'intelligibilité en association avec les aspects stylistiques : débit, spontanéité et clarté de la prononciation. Cette conception est formulée par la théorie de la *variabilité adaptative* (ou de l'*hyper/hypoarticulation*) de Lindblom (1990). Reste que la coarticulation n'a pas forcément pour effet de diminuer l'information portée par les phonèmes, mais constitue aussi d'un autre côté une manœuvre pour améliorer ou préserver la robustesse du signal. En effet, en distribuant une propriété phonétique d'un son sur une portion plus grande du signal, c'est-à-dire sur les segments voisins, la redondance de l'information linguistique est accrue. Elle est alors moins fragile face aux détériorations de la qualité du signal (brièveté, bruit masquant, faible intensité…) et peut être ainsi plus aisément retrouvée, même si un segment donné est inaudible ou inintelligible (Wright 2004). La coarticulation peut donc être considérée également comme un support à l'intelligibilité de la parole, jouant un rôle dans la reconnaissance des mots, en facilitant et accélérant l'accès à leur forme sonore (Nguyen 2001).

Particulièrement centrale dans les processus de production de la parole, la prise en compte des phénomènes de coarticulation est à l'origine d'influents modèles de production, telle la *dynamique de tâches* de Saltzman (1986), et phonologiques, telle la

phonologie articulatoire de Browman & Goldstein (1992)[16], pour lesquels les gestes articulatoires constituent les unités primitives d'action et de représentation sonores, allant jusqu'à postuler l'inexistence du phonème. Mais, la coarticulation joue également un rôle décisif dans les processus de perception. Ainsi, elle est au cœur de divers modèles de perception de la parole, comme la *théorie motrice*[17] de Liberman & Mattingly (1985), et ses variantes (notamment Fowler 1986). De même, la *théorie perceptive du changement phonétique* d'Ohala (1981) place la perception de la coarticulation comme un moteur essentiel du processus diachronique de mutations phonologiques des langues : les phonèmes d'une langue évoluent principalement sur la base d'une perception sous-tendue par les connaissances intériorisées des locuteurs sur les faits systématiques de coarticulation. La coarticulation apparaît donc comme capitale pour appréhender le fonctionnement tant de la parole que de la langue.

Cependant, les différentes unités sonores ne sont pas égales face à ce phénomène, comme le modèle de coarticulation du *degré de contrainte articulatoire* de Recasens & Espinosa (2009) propose d'en rendre compte concernant l'articulation linguale. Les travaux de Recasens et ses collègues sur la coarticulation entre consonnes et voyelles en catalan confirment les résultats obtenus sur d'autres langues (Farnetani 1997). Ils montrent que le degré de résistance d'un segment à l'influence de ses voisins est plus grand, plus l'articulateur a une inertie mécanique importante et plus il est massivement impliqué dans le geste de constriction. Articulateur, lieu et mode d'articulation constituent donc des contraintes qui conditionnent le degré de coarticulation entre segments. Plus une articulation implique une large surface de contact linguopalatal (notamment du dos de la langue), par une portion importante de la langue réalisant une aire de constriction étendue, plus cette articulation est résistante, ne laissant qu'une petite partie linguale libre d'anticiper ou de conserver une position proche de celle adoptée pour les segments suivants ou précédents. Ainsi, l'articulation d'une consonne palatale, comme [ɲ], ou alvéo-palatale, comme [ʃ], impliquant une partie conséquente du dos de la langue, voire de la lame, se montre moins variable en fonction du contexte vocalique qu'une consonne alvéolaire telle que [n], [l] ou [t], et encore bien moins qu'une labiale comme [p] où la langue reste totalement disponible. De même, les fricatives présentent un mode d'articulation particulièrement contraignant par la réalisation d'un chenal fricatif mobilisant une grande partie de la langue. La fricative alvéo-palatale [ʃ] est ainsi l'une des consonnes les plus résistantes à la coarticulation linguale. Similairement [i], par son articulation fermée et pré-palatale impliquant fortement le corps de la langue, est la voyelle la plus résistante, bien plus que [a] et [u], dont la constriction moins étroite ou étendue recrute moins le corps lingual, libérant largement sa majeure partie antérieure. La coarticulation affecte donc principalement les régions ne participant pas à la constriction, mais dont un changement peut plus ou moins impacter la forme spectrale du son. Parallèlement, résistance et influence coarticulatoires sont positivement corrélées. Cela signifie que plus un segment est

[16] Pour une présentation récente et détaillée, voir Fougeron (2005).

[17] Pour une revue critique récente, voir Galantucci et al. (2006).

insensible aux effets de la coarticulation, plus il exerce une influence sur l'articulation des autres segments. Par exemple, Niebuhr et al. (2008) montrent en français que les sibilantes [s z] sont l'objet d'une palatalisation au contact de [ʃ ʒ], quelle que soit leur place dans les séquences (alvéolaire-postalvéolaire ou postalvéolaire-alvéloaire), indiquant que les fricatives postalvéolaires exercent une attraction dominante sur le lieu alvéolaire. Toutefois, la coarticulation concerne tous les articulateurs, et donc si une articulation peut contraindre fortement la position de la langue, celle des autres articulateurs non impliqués directement (par exemple, les lèvres ou le velum) peut plus librement s'accorder aux configurations des segments voisins. Ainsi, les lèvres adoptent systématiquement une configuration arrondie lors d'un [ʃ] suivi de [y] et étirée avant [i], ce qui en modifie le timbre. Reste que d'autres forces s'appliquent également sur les processus de coarticulation, notamment des contraintes biomécaniques, aérodynamiques, perceptives et de synchronisation des gestes articulatoires. Par exemple, il est observé dans diverses langues (Bombien et al. 2010), dont le français (Marchal 1988), un effet d'ordre des lieux d'articulation sur le timing gestuel dans les groupes de consonnes : l'enchaînement postéro-antérieur (de type [kt] ou [kp]) montre un chevauchement articulatoire, à savoir une anticipation du geste occlusif de la seconde consonne, moins important qu'une séquence antéro-postérieure (de type [tk] ou [pk]). Ces synchronisations différentes reposeraient sur une motivation acoustico-perceptive. L'occlusion antérieure réalisée avant le relâchement de la consonne postérieure précédente en masquerait le bruit d'explosion, alors que l'explosion d'une consonne antérieure reste audible même si l'occlusion de la consonne postérieure suivante est effectuée avant. Néanmoins, toutes les conséquences spectrales du timing articulatoire restent difficiles à appréhender, ces faits coarticulatoires n'étant pas toujours accessibles en surface du signal acoustique. Une autre raison à cela est que les processus de coarticulation sont essentiellement graduels et non catégoriels dans leur réalisation, tant dans la dimension temporelle que spatiale (Holst & Nolan 1995 par exemple). Les gestes se chevauchent, s'influencent ou fusionnent plus ou moins et de manière complexe en fonction de multiples facteurs : (i) biomécaniques (cf. ci-dessus) ; (ii) linguistiques, relatifs à la nature phonétique, à leurs position dans les unités syllabique, lexicale ou prosodiques (cf. ci-après) et aux contraintes spécifique de la 'grammaire' de chaque langue (Farnetani 1997) ; (iii) extralinguistiques, relatifs aux impératifs de la situation de communication (cf. ci-avant).

Avant de poursuivre sur les phénomènes coarticulatoires communs en français spontané, il semble nécessaire d'éclaircir les termes connexes de 'coproduction', 'coarticulation', et 'assimilation' (Farnetani 1997). La notion de coproduction n'est pas strictement synonyme de coarticulation, mais est plus générale. Elle renvoie plus précisément au processus biomécanique intrinsèque à l'acte moteur impliquant la production simultanée ou chevauchée de mouvements articulatoires coordonnés réalisant les sons enchaînés dans la parole. La coarticulation n'en est en fait qu'une conséquence phonétique tant spatiale (attraction d'une position articulatoire sur une autre) que temporelle (synchronisation des gestes). Le mot d'assimilation est généralement plus restrictif, réservé aux phénomènes de coarticulation spatiale affectant le lieu, le degré d'ouverture (mode consonantique, aperture vocalique, latéralité ou nasalité) et le voisement des sons, rendant les sons voisins plus similaires.

Traditionnellement, ce terme était plus restreint à un point de vue perceptif ou phonologique, dans ce cas rattaché strictement aux faits d'allophonie phonologisée (à savoir binaire ou catégorielle). Arrêtons-nous à présent sur quelques cas de coarticulation en français, opérant communément en parole spontanée et dont les effets acoustiques sont les plus aisément observables sur le signal. Comme dans toute langue, les faits de coarticulation (et de réduction) sont plus fréquents, variés et marqués, plus la parole est spontanée et rapide (Duez 2001). Ces phénomènes de coarticulation concernent un ensemble d'assimilations (i) de lieu d'articulation : antériorisation, postériorisation, labialisation, palatalisation, vélarisation, (ii) de mode d'articulation : degré de constriction, aperture, nasalisation, (iii) et de voisement : assourdissement, sonorisation, affectant voyelles et consonnes. Les assimilations ont lieu dans deux directions alternatives ou concomitantes : régressive quand un son est influencé par anticipation d'un son suivant, ou progressive quand un son subit l'effet de persévération d'un son précédent. Mais la systématicité de ces directions reste malaisée à déterminer. On sait malgré tout que certains facteurs la conditionnent, par exemple la nature de l'articulation, comme on a pu le voir s'agissant de l'articulation alvéo-palatale [ʃ] attractive quelle que soit sa position. Un autre grand facteur connu est justement la position lexicale et syllabique. La position finale de mot ou de syllabe rend le son plus perméable aux phénomènes de coarticulation (et de réduction) que la position initiale, considérée comme une position métriquement forte. C'est d'ailleurs cette position finale qui supporte les principales modifications phonétiques diachroniques, par exemple du latin au français (Straka 1964). Ainsi, très fréquemment une consonne finale par exemple se dévoise ou se voise ([vɔtka] et non *[vɔdga] pour « vodka », [sag də sabl] et non *[sak tə sabl] pour « sac de sable »), se nasalise ([manmwazɛl] et non *[mãdmwazɛl] pour « mademoiselle », [ʁan militɛʁ] et non *[ʁad mɪ̃litɛʁ] pour « rade militaire »), se postériorise ([aŋ fɐbl] et non *[am ɸɛbl] pour « âme faible ») ou s'antériorise ([pɔβ mɛk] et non *[pɔv ŋɛk] pour « pauvre mec »)[18]. La coarticulation anticipatoire est communément rapportée comme plus fréquente, notamment en français (Carton 1974, Duez 2001) ; en témoignent les modèles de production articulatoire focalisés sur l'anticipation plus que sur la persévération articulatoire (Farnetani 1997). L'assimilation de lieu des consonnes est un phénomène de coarticulation anticipatoire largement récurrent. Si l'assimilation de lieu est décrite comme faible dans les groupes de consonnes en français (Duez 2001, 2003), celle entre une consonne et la voyelle suivante est plus systématique, bien que de différents degrés selon les contextes. Particulièrement, les consonnes se palatalisent au contact d'une voyelle fermée et antérieure, c'est-à-dire que leur articulation principale se déplace légèrement, ou bien qu'une constriction secondaire de la lame de la langue est opérée, dans la zone palatale de [i] ou [j]. Ainsi, les consonnes coronales, comme [t d n], se postériorisent et les consonnes postérieures vélaires [k g] ou uvulaires [ʁ χ] s'antériorisent avant [i y e][19].

---

[18] [ŋ] est une consonne nasale labio-dentale de même lieu que [f v], plus arrière que les bilabiales nasale [m] ou fricatives voisée [β] et sourde [ɸ]. Pour une présentation détaillée des signes de l'Alphabet Phonétique International, voir notamment Durand (2005).

[19] Ce fait de coarticulation s'est phonétiquement renforcé, lexicalement fixé et phonologiquement diffusé en français méridional où le pronom « tu » [ty] est réalisé [tʲ] quelle que soit la voyelle qui

Acoustiquement, la palatalisation se traduit lors du relâchement de la consonne par l'apparition d'un pic d'énergie autour de 2000 à 2500 Hz dans le spectre du bruit d'explosion et de friction, associé au F_2 haut de la voyelle antérieure (Ladefoged & Maddieson 1996). Également, combinée à la palatalisation, les voyelles antérieures fermées provoquent mécaniquement une affrication (ou spirantisation) des plosives, c'est-à-dire une augmentation de l'intensité et de la durée du bruit suivant l'explosion de la consonne. Ce phénomène est physiologiquement conditionné par le maintien, après l'explosion, de l'avant de la langue dans une position très fermée (celle produite pour [i]), ce qui a pour effet de conserver plus longtemps une PIO importante génératrice de turbulences aérodynamiques. Ainsi, le [t] de [ti] se rapproche acoustiquement et perceptivement de [tˢi] ou [tʲⁱi], alors que suivi d'une voyelle plus ouverte comme dans [ta], ce bruit est quasi absent, du fait qu'après le relâchement de [t] l'avant de la langue atteint rapidement une position très basse pour l'articulation de la voyelle ouverte, empêchant le maintien d'une PIO élevée[20]. La labialisation des consonnes avant voyelle arrondie est également un cas systématique d'anticipation articulatoire où la protusion des lèvres est initiée bien avant le geste lingual réalisé pour la voyelle. Ainsi, Benguerel & Cowan (1974) ont montré pour le français que le geste d'arrondissement des lèvres peut remonter la chaîne de parole jusqu'à la voyelle précédente impliquant un geste antagoniste d'étirement des lèvres, comme pour [i], quel que soit le nombre de consonnes intervocaliques. Du fait de l'allongement du conduit vocal par la protusion labiale, la labialisation a pour effet acoustique d'abaisser les fréquences des bruits d'explosion et de friction des obstruantes (figure 1.10, [s] de [se] *vs* [s] de [sy]) et des formants des sonantes. Un autre phénomène d'anticipation articulatoire à distance concerne l'harmonisation vocalique. Loin d'être un processus phonologique comme en hongrois, turc, finnois ou shona, en français il concerne les voyelles moyennes fermées/ouvertes [e/ɛ] [o/ɔ], voire [œ/ø], en syllabe ouverte pénultième de mot. Celles-ci tendent à adopter l'aperture de la voyelle accentuée finale de mot : par exemple « aimer » tend à être produit [eme], alors que « aimable » [ɛmabl], de même pour « moto » [moto] *vs* « moteur » [mɔtœʁ] et « gueuler » [gøle] *vs* « (ils) gueulèrent » [gœlɛʁ]. Bien que les descriptions de ce phénomène varient selon les auteurs[21], il est communément dépeint comme optionnel, variable selon les régions et les locuteurs, et plus fréquent en parole informelle. Nguyen & Fagyal (2008) fournissent la première validation acoustique expérimentale de ce phénomène en français. Leur étude met en évidence que le F_1 des voyelles moyennes est plus haut si elles sont suivies par une voyelle mi-fermée ou fermée en syllabe finale de mot. En outre, F_2 est plus élevé lors

---

suit : « tu es » [tʲe], « tu as » [tʲa], « tu auras » [tʲoʁa], « tu ouvres » [tʲuvʁ], « tu en as » [tʲɑ̃na], etc. Cette prononciation est par contre impossible, si « tu » n'est pas pronom. Par exemple, « tuer » [tye] ou [tɥe] et non *[tʲe], ou « (il) tua » [tya] ou [tɥa] mais non *[tʲa]. Ce phénomène constitue un marqueur phonétique majeur de ce régiolecte.

[20] Ce phénomène a donné lieu à une phonologisation en allophones non affriqués *vs* affriqués des plosives alvéolaires /t d/ en français québécois et marseillais. Avant [i y], /t d/ sont réalisés en consonnes affriquées [t͡s d͡z] en québécois et [t͡ʃ d͡ʒ] (accompagné d'une palatalisation) en marseillais 'des quartiers'.

[21] Voir Nguyen & Fagyal (2008) pour une revue détaillée.

que la voyelle moyenne précède [i] (*vs* [a]), attestant qu'un déplacement du lieu d'articulation est également possible. Ce phénomène est plus marqué pour les locuteurs du Nord que du Sud de la France et reste un phénomène optionnel et graduel de coarticulation, et non catégoriel (allophonique). Des études manquent cependant pour vérifier le fait que l'harmonisation vocalique en français n'est pas restreinte à ce contexte donné, mais peut aussi concerner des syllabes fermées[22] (par exemple, la prononciation en [e] pour « vestige » ou « technique » *vs* en [ɛ] pour « vestale » ou « secteur »), et remonter plus loin dans le mot (par exemple [pʁofɛsjɔnɛl] ou [pʁɔfɛsjɔnɛl] pour « professionnel »). Assimilation tant anticipatoire (ou régressive) que persévérative (ou progressive) existent concernant la nasalisation, posant les segments nasals comme une propriété articulatoire attractive. Au contact d'une nasale, une consonne ou voyelle orale tend plus ou moins à se nasaliser, à l'inverse d'une nasale qui ne se dénasalise que très rarement. Le contexte le plus favorable à la nasalisation est celui d'une voyelle précédée et suivie d'une consonne nasale, par exemple « même » /mɛm/ produit [mɛ̃m]. La nasalisation progressive est cependant bien plus fréquente et importante en français. Un segment oral est bien plus souvent et largement nasalisé, partiellement ou totalement, si le segment nasal précède que s'il suit. Ainsi, les études aérodynamiques de Basset et al. (2001) et Delvaux (2000, 2003) montrent que l'anticipation du geste d'abaissement du voile reste faible pour les consonnes précédant une voyelle nasale, d'autant plus que leur mode d'articulation est fermé : plosives < fricatives < liquides. Ce faible chevauchement de l'abaissement du velum avec l'articulation des plosives permettrait de préserver leur bruit d'explosion constituant un indice acoustique important pour leur identification perceptive. De même, une voyelle avant une consonne nasale est peu voire pas nasalisée. A l'inverse, une voyelle ou une consonne après nasale est l'objet d'une forte nasalisation, du fait d'une latence importante du geste de fermeture du voile. Ces données confirment la plupart des résultats aérodynamiques de Cohn (1993) et des observations acoustico-perceptives de Duez (2003) sur les groupes de consonnes en parole conversationnelle. Duez (2001) précise que les consonnes voisées sont plus sensibles que les sourdes à l'assimilation de nasalité. Les conséquences acoustiques de la nasalisation sont celles décrites ci-avant, relatives à la nasalité vocalique (cf. §1.3.1) et consonantique (cf. §1.3.2). Enfin, un autre cas de coarticulation persévérative concerne le dévoisement après obstruantes sourdes. Il est largement observé que les consonnes liquides /l ʁ/[23] se dévoisent au moins partiellement après une plosive ou une fricative sourde en français (Meunier 1994), comme dans « plus » [pl̥y] (figure 1.11). Ce phénomène pourrait être mis au compte de la plus grande cohésion syllabique de ce type de groupe, obstruante-liquide /pl bl pʁ bʁ tʁ dʁ kl gl kʁ gʁ fl fʁ vʁ/, qui est toujours tautosyllabiques, alors que tout autre groupe de consonnes en français

[22] Une syllabe ouverte est une syllabe qui se termine par une voyelle prononcée, et une syllabe fermée par une consonne.

[23] /ʁ/ est considéré ici comme une liquide (et non une fricative) du fait de sa forme historique /ɾ/ (sonante) et que phonétiquement elle présente souvent une forme affaiblie où les composantes de bruit sont réduites ([ɻ]).

est hétérosyllabique[24]. Ce phénomène apparaît néanmoins aussi pour des types de groupes de consonnes différents comme dans « cheveu » prononcé [ʃy̥ø] (après chute du schwa), « snob » [sn̥ɔb] ou smala [sm̥ala]. Le dévoisement semble donc une question de prédominance de la position initiale sur la position non initiale de syllabe plutôt que de cohésion syllabique. En parole rapide ou spontanée, le dévoisement persévératif peut aussi affecter les voyelles, notamment fermées antérieures après obstruantes sourdes. Ainsi, les dévoisements partiels vocaliques sont fréquents : [y] de « sûr » (figure 1.10), [e] de « sais » (figure 1.11) ou de « effectivement » (figure 1.12) ; aussi, les dévoisements complets ne sont pas rares, comme par exemple le [i] de « effectivement » (figure 1.12). Ces cas de dévoisement des voyelles fermées semble aérodynamiquement conditionné par le fait qu'au relâchement de la consonne la langue reste en position très fermée participe à maintenir plus longtemps une PIO élevée, retardant l'initiation vibratoire des cordes vocales du fait d'une différence de pression trop faible entre PSG et PIO. Enfin, un cas plus particulier d'assimilation de voisement affecte certains proclitiques en français : « je », « le », « se », « ne », « de » parmi d'autres. Ces mots grammaticaux sont des morphèmes libres métriquement faibles, du fait de leur caractère non accentogène et de leur composition syllabique instable. Ceux sont des mots principalement monosyllabiques dont le noyau vocalique est communément un schwa [ə], sujet à des élisions fréquentes. Ils se lient toujours au mot suivant (adjectif, nom, verbe…) pour former un groupe prosodique minimal accentué sur la syllabe finale. Après élision du schwa, la consonne initiale restante du clitique s'assimile toujours en voisement à la consonne initiale du mot suivant auquel le clitique est lié. Ainsi, en parole spontanée on trouvera par exemple [ʃse] (figure 1.11) ou [ʃːe] pour « je sais » et jamais *[ʒse] ou *[ʒze], ou encore [i zvwa] pour « il se voit » et non *[i svwa] ou *[i sfwa]. Bien qu'étant une consonne initiale de mot, plus résistante à la coarticulation, le caractère faible des clitiques semble conditionner son assimilation régressive. Par contre, le fait que dans « je sais » [s], pourtant initial du mot lexical, puisse s'assimiler au lieu d'articulation de la consonne du clitique « j(e) » précédent semble relever de la dominance articulatoire de la consonne alvéo-palatale [ʃ] plus résistante et attractive par rapport à l'alvéolaire (cf. ci-avant).

Le concept de réduction peut être vu en partie comme une autre conséquence phonétique de la coproduction, distinguable des faits de coarticulation, même s'ils sont souvent combinés ou confondus. La réduction est un affaiblissement phonétique des phonèmes, syntagmatiquement moins contrastifs et/ou paradigmatiquement moins distinctifs, c'est-à-dire sous-articulés. Elle renvoie à une diminution de l'amplitude et/ou de la durée des déplacements gestuels, et plus globalement de l'effort articulatoire. Sous l'angle plus particulier de la coproduction, deux mécanismes principaux peuvent conduire à une réduction acoustique des sons dans la chaîne parlée : la troncation et le masquage. La troncation consiste en un blocage du geste ou mouvement articulatoire au

[24] Deux consonnes tautosyllabiques appartiennent à la même syllabe : par exemple, [a.gʁym] « agrume » ou [o.bl̥ik] « oblique » ; deux consonnes hétérosyllabiques sont séparées par une frontière de syllabe (.) : par exemple, [aʁ.gys] « argus » ou [ɔp.tik] « optique ».

profit du suivant. Elle peut conduire à sa diminution spatiale (lieu ou ouverture) et/ou son raccourcissement temporel (Beckman & Cohen 1999). Ce phénomène apparaît surtout dans le cas d'enchaînement de gestes antagonistes homo-organiques (impliquant le même articulateur). Par exemple, le fait largement constaté qu'une voyelle est plus courte devant une consonne non voisée qu'une voisée peut s'expliquer par la troncation du geste d'adduction (accolement) des cordes vocales, réalisé pour le voisement de la voyelle, par le geste antagoniste d'abduction (écartement), nécessité par la consonne sourde suivante. Le masquage d'un son, ou de l'une de ces propriétés, a également pour conséquence une réduction temporelle ou spectrale. Sur le plan articulatoire, il consiste en un chevauchement important de gestes atténuant ou occultant un trait acoustique pourtant articulatoirement produit. Par exemple dans un mot comme « gnou », l'abaissement du velum pour [n] avant le relâchement de [g] a pour effet acoustique d'affaiblir considérablement ou de masquer totalement le bruit d'explosion de la plosive vélaire, même si son occlusion est effectivement relâchée. Reste que la réduction ne sous-entend pas nécessairement une influence phonétique des sons contigus, mais est le produit d'une articulation plus relâchée où les cibles linguistiques sont approchées et non atteintes. Le fait de ne pas atteindre chaque cible articulatoire consécutive dans la chaîne parlée entraîne une diminution du contraste syntagmatique entre les sons, ce qui peut-être équivalent en surface à considérer que chaque cible successive se rapproche l'une de l'autre, et que donc la coarticulation opère davantage. Ainsi, si l'on prend le cas de consonnes fricatives s'affaiblissant en approximante entre deux voyelles, comme /ʁ/ en français, et inversement celui d'une moins grande aperture des voyelles ouvertes entre deux consonnes, réduction et coarticulation (assimilation de mode/aperture) se confondent. Il existe donc bien un lien étroit entre sous-articuler et coarticuler, mais sans identité stricte de ces phénomènes. En effet, parler moins fort ou de façon plus relâchée entraîne une réduction généralisée de l'effort et des déplacements articulatoires, mais n'inclut pas de considérer qu'un accroissement de l'influence spécifique des contextes phonétiques locaux en est la cause directe. Reste que, répondant parallèlement aux mêmes nombreux facteurs phonétiques, linguistiques, stylistiques et communicationnels (cf. ci-avant ; Duez 2001), coarticulation et réduction sont tous deux tellement intriquées dans l'acte de production que leur imbrication est par nature souvent inextricable. Leur proximité étroite paraît particulièrement s'exprimer dans les métaplasmes observés en parole spontanée (Kohler 1998 pour l'allemand, Johnson 2004 pour l'anglais, Adda-Decker et al. 2008, ce volume pour le français). Ces objets sonores apparaissent comme des contractions phonétiques extrêmes de mots ou d'expressions plus ou moins amalgamées et/ou tronquées, lieu privilégié d'élision, de modifications, voire de fusions articulatoires complexes, mêlant très étroitement des mécanismes avancés de réduction (allant souvent jusqu'à l'élision) et de coarticulation. Les figures 1.12 et 1.13 présentent des métaplasmes en parole spontanée montrant respectivement une forme phonétique assez et extrêmement contractée, dont il peut être plus ou moins difficile de donner une transcription précise. Johnson (2004) montre qu'en anglais les 'réductions massives' (selon ses termes) sont loin d'être anecdotiques en parole conversationnelle. Ainsi, il relève que 10 à 20% des mots plurisyllabiques sont produits avec au moins une syllabe ou un noyau syllabique élidé, qu'entre 20 à 40% présentent une modification phonétique de leur forme sonore 'canonique', et que plus le mot est long et fonctionnel (*vs* lexical), plus il est susceptible de réduction. Globalement un mot sur vingt présente une réduction

par élision ou modification phonétique conséquente. La compréhension de la reconnaissance perceptive de ces formes est un enjeu nouveau pour les modèles de traitement de la parole, laissant entrevoir que les processus 'top-down' (des représentions de haut niveau : sémantique, pragmatique, syntaxique, lexicale, etc, aux signaux phonétiques…) ne suffisent pas à expliquer leur perception, et que probablement des processus 'bottom-up' (du signal au message) importants sont également en action. Ainsi, Meunier (communication personnelle) ou Johnson (2004), évoquant respectivement un 'socle articulatoire' et un 'ilot de fiabilité' pour ces cas de réductions extrêmes, pensent des conditions phonétiques nécessaires à leur assise perceptive. Outre ces cas, nous terminerons sur des phénomènes de réduction communs et fréquemment relevés en parole spontanée en français. L'élision est une forme ultime de réduction qui est loin d'être rare en conversation. Voyelles et consonnes sont susceptibles d'être omises, même si en position métriquement faible finale de syllabe ou de mot et inaccentuée, elles le sont davantage. Typiquement, avec les sonantes nasales, les liquides [l ʁ] sont les meilleurs candidats à l'élision notamment en finale (Duez 2003), ainsi par exemple les prononciations communes [pɔv ɡa] pour « pauvre gars » et [a tab] pour « à table », ou encore le [ʁ] de « sûr » dans la figure 1.10. Mais les obstruantes en sont également sujettes comme l'indique la prononciation [ɛfet̪ʃi̥y̥ma͡ɑ̃][25] du mot « effectivement » (figure 1.12) qui révèle une élision complète de la plosive /k/ de la seconde syllabe et un fort affaiblissement de la fricative /v/ de la troisième. Cet exemple illustre donc également le phénomène d'affaiblissement des consonnes. L'affaiblissement des consonnes reposent principalement sur le voisement ou dévoisement des consonnes et sur une ouverture plus grande de leur mode d'articulation : les plosives étant produites comme des fricatives ou affriquées, et les fricatives en approximantes (Duez 2001 entre autres). Ainsi, dans ce même exemple, outre le dévoisement avancé et le fort raccourcissement de /v/, le /t/ se spirantise en [t̪ʃ], c'est-à-dire est produit avec un mode d'articulation plus ouvert, provoquant une absence d'occlusion complète et donc d'explosion acoustique au profit d'un bruit de friction plus long. Outre leur assourdissement (cf. ci-avant), les voyelles sont également fréquemment susceptibles de réduction tant dans le domaine spatial que temporel. De nombreuses études en français (Meunier & Espesser, 2011, Adda-Decker et al., ce volume), et dans d'autres langues (Gendrot & Adda-Decker 2007), confirmant les corrélations entre durée et timbre vocalique connues depuis les travaux pionniers de Straka (1963) et Lindblom (1963), montrent qu'une réduction des propriétés articulatoires et spectrales des voyelles accompagne la diminution de leur durée. Ainsi, la figure 1.5 présente la rétraction du triangle vocalique des voyelles en fonction du style de parole (lecture *vs* conversation spontanée). Des voyelles lues aux voyelles spontanées et des plus longues aux plus courtes, un phénomène de centralisation croissante des voyelles est mis en évidence, plus le discours est relâché, plus le débit est rapide, et/ou plus les voyelles occupent des positions métriquement faibles, à savoir médiane de syllabe ou de mot, et prosodiquement éloignées des frontières intonatives, lieux d'allongement syllabique (cf. ci-après). Dans ces contextes de la parole conversationnelle, celles-ci perdent une partie de leur caractéristiques contrastives, les

[25] A noter la prononciation méridionale de la voyelle nasale finale /ɑ̃/.

rendant moins différentes, et donc moins périphériques dans l'espace vocalique. Parmi les dimensions articulatoires des voyelles, la plus forte variation du F_1 indique que l'ouverture du conduit (aperture) est particulièrement affectée lors des réductions vocaliques en parole spontanée ; les voyelles plus ouvertes présentent des variations beaucoup plus importantes que les autres. Le schwa, ou « e muet », /ə/ est un autre grand domaine familier d'application de la réduction phonétique en français. Son élision, très commune dans certains mots et contextes (voir par exemple Tranel 1987, Racine & Grosjean 2002), est présentée dans nombre d'études (notamment Lucci 1983, Hansen 1994) comme un marqueur du français parlé (discours, interview) face à l'écrit oralisé (lecture). Celle-ci pourrait être vue comme une sorte de 'degré zéro' de la réduction en prononciation spontanée. Deux remarques paraissent intéressantes concernant ce phénomène en parole conversationnelle. Tout d'abord, l'étude phonétique Bürki et al. (2007) montre sur un grand corpus radiophonique que l'élision du schwa est dans près de 10% des cas perçue comme ambiguë. Les corrélats acoustiques de cette ambiguïté ne sont pas claires, mais la durée acoustique de l'élément est corrélée à sa perception : plus il est long plus, plus il est perçu comme présent ; et plus il est court, plus sa perception est ambiguë ou est jugé absent. Il en ressort qu'il est possible que l'effacement de schwa puisse relever d'un processus graduel et non discret, même si un conditionnement articulatoire lié au contexte consonantique semble opérer. Par ailleurs, on peut également noter que l'effacement du schwa pourrait constituer une certaine condition pour des réductions ou des modifications phonétiques du mot plus importantes. En effet, par exemple, sans son élision le [v] de « cheveu » ne pourrait pas se dévoiser sous l'influence coarticulatoire de la consonne sourde initiale de mot [ʃ] : [ʃəvø] ou [ʃv̥ø], mais pas *[ʃəv̥ø] ou *[ʃvø]. De même, la prononciation d'« effectivement » en [ɛfɛt̪ˢi̯v̥mɑ̃͡ɑ̃] (figure 1.12), à savoir avec élision de [və], n'apparaît possible que si l'élision presque complète de [ə] est réalisée. La chute du noyau vocalique semblant pouvoir conditionner la chute totale de la syllabe, et donc de la consonne initiale ; sans l'élision du [ə], [v] ne pourrait très probablement pas s'affaiblir, voire tomber. Ainsi, on peut supposer qu'un ordre dans le degré de réduction phonétique des mots opère : par exemple pour la syllabe /və/, on pourrait postuler les étapes de réduction croissante suivante : [ɛfɛktivəmɑ̃] > [ɛfɛktivmɑ̃] > [ɛfɛkt̪i̯v̥mɑ̃] > [ɛfɛkt̪i̯mɑ̃] > etc. L'élision du schwa pourrait donc être une condition nécessaire, mais non suffisante, à une sorte de réaction en chaîne vers des formes plus coarticulées et réduites.

Enfin, il reste à préciser que si les phénomènes de réduction et de coarticulation sont fréquents et récurrents en parole spontanée et conversationnelle, cela ne signifie pas qu'ils opèrent aveuglément et à tous les points de la chaîne de parole. Comme on l'a vu, des facteurs stylistiques (débit, spontanéité, exigence d'intelligibilité), contextuels (environnement phonétique) et positionnels (face à la structure syllabique, lexicale ou morphologique) sont en œuvre. Un de ces facteurs tout aussi important est relatif à la position du mot ou du segment au sein de la structure prosodique des énoncés. L'accentuation et l'intonation participent à mettre en relief certains mots ou syllabes et à organiser les messages en regroupements de mots à différents niveaux hiérarchiques : les syntagmes ou constituants prosodiques, tels que le mot prosodique, le groupe accentuel ou le groupe intonatif. Nombre d'études, dans de multiples langues (Keating

et al. 2003 par exemple), dont le français, ont montré que les segments produits aux points clefs de cette structure prosodique sont moins coarticulés (Meynadier 2010) et voit généralement leur articulation renforcée (Fougeron 2001). A savoir sous un accent et/ou en frontière intonative (avant ou après, c'est-à-dire en position finale ou initiale de constituant prosodique), ils tendent à mieux maintenir une articulation plus contrastive avec les segments contigus de la chaîne parlée et des caractéristiques plus distinctives avec les autres éléments du système phonologique. Ainsi, la prosodie est un acteur essentiel non seulement de l'organisation accentuelle et intonative des messages sonores, mais aussi des processus de production articulatoire et acoustique des segments intégrant les énoncés de parole**.**

1.5. Bibliographie

Adda-Decker M., Gendrot C. & N. Nguyen (ce volume). Apport du traitement automatique à l'étude des voyelles. In N. Nguyen (éd.), *Méthodes et outils pour l'analyse phonétique des grands corpus oraux*. Paris : Hermès.

Amelot A. (2004). *Étude aérodynamique, fibroscopique, acoustique et perceptive des voyelles nasales du français*. Thèse de doctorat. Université de la Sorbonne Nouvelle, Paris.

Auran C., Espesser R., Goldman J.-P. & N. Nguyen (ce volume). Le traitement et l'analyse du signal de parole. In N. Nguyen (éd.), *Méthodes et outils pour l'analyse phonétique des grands corpus oraux*. Paris : Hermès.

Badin, P. (1991). Fricatives consonants: acoustic and X-ray measurements. *Journal of Phonetics*, 19, 397-408.

Basset P., Amelot A., Vaissière J. & B. Roubeau (2001). Nasal flow in French Spontaneous Speech. *Journal of the International Phonetic Association*, 31, 87-100

Beckman M. & K. B. Cohen (1999). Modeling the articulatory dynamics of two levels of stress contrast. In M. Horne (ed.), *Prosody: Theory and Experiment*, pp. 169-200. Dordrecht : Kluwer Academic.

Benguerel A. & H.A. Cowan (1974). Coarticulation of upper lip protusion in French. *Phonetica*, 30: 41-55.

Bertrand R., Blache P., Espesser R., Ferré G., Meunier C., Priego-Valverde B. & S. Rauzy (2008). Annotation et exploitation multimodale de parole conversationnelle. Traitement Automatique des Langues, 49, 1-30.

Bombien L., Mooshammer C., Hoole P. & B. Kühnert (2010). Prosodic and segmental effects on EPG contact patterns of word-initial German clusters. *Journal of Phonetics*, 38, 388-403.

Bothorel A., Simon P., Wioland F. & J-P. Zerling (1986). *Cinéradiographie des voyelles et des consonnes du français*. Strasbourg : Institut de Phonétique.

Browman C. P. & L. Goldstein (1992). Articulatory Phonology: An Overview. *Phonetica*, 49, 155-180.

Bürki A., Fougeron C., Gendrot C. & U. Frauenfelder (2007). De l'ambiguïté de la chute du schwa en français. In *Actes des 5è Journées d'Etudes Linguistiques*, 83-88. Nantes.

Carton F. (1974). *Introduction à la phonétique du français.* Paris : Bordas.

Cohn A.C. (1993). Nasalisation in English: phonology or phonetics ? *Phonology*, 10: 42-81.

Coveney A. (2001). *The Sounds of Contemporary French.* Exeter: Elm Bank.

Delattre P., Liberman A. M. & F. S. Cooper (1955). Acoustic loci and transitional cues for consonants. *Journal of Acoustical Society of America*, 27, 769-773.

Delvaux V. (2000). Etude aérodynamique de la nasalité en français. *Actes des XXIIIèmes Journées d'Etude sur la Parole*, pp. 141-144. Aussois.

Delvaux V. (2003). *Contrôle et connaissance phonétique : les voyelles nasales du français*. Thèse de doctorat. Université Libre de Bruxelles.

Delvaux V., Metens T. & A. Soquet (2002). Propriétés acoustiques et articulatoires des voyelles nasales du français. *Actes des 24è Journées d'Etude sur la Parole*, pp. 348-352. Nancy.

Duez D. (2001). Manifestation phonétique de la réduction et de l'assimilation contextuelle des segments de la parole conversationnelle. *Revue PArole*, 17-18-19 : 89-111.

Duez D. (2003). Acoustic properties of consonant sequences in conversational French speech. *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 2965-2968. Barcelona.

Durand J. (2005). La phonétique classique : l'Association Phonétique Internationale et son alphabet. In N. Nguyen, S. Wauquier-Gravelines & J. Durand (dir.), *Phonologie et phonétique : forme et substance*, pp. 25-60. Paris : Hermès.

Fant G. (1960). Acoustic *Theory of Speech Production*. La Haye: Mouton.

Farnetani E. (1997). Coarticulation and connected speech processes. In W. J. Hardcastle & J. Laver, *Handbook of Phonetic Sciences*, pp. 371-404. Oxford: Blackwell.

Fougeron C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, 29, 109–135.

Fougeron C. (2005). Introduction à la Phonologie Articulatoire. In N. Nguyen, S. Wauquier-Gravelines & J. Durand (dir.), *Phonologie et Phonétique*, pp. 265-290. Paris : Hermès.

Fougeron C. & C. Smith (1999). French. In J. H. Esling (ed.), *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*, pp. 78-81. Cambridge: Cambridge University Press.

Fowler C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3-28.

Galantucci B., Fowler C.A. & M. T. Turvey (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13, 361-377.

Gendrot C. & M. Adda-Decker (2007). Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages. *Proceeding of the International Conference of Phonetic Sciences*, pp. 1417-1420. Saarbrücken.

Ghio A. & S. Pinto (2007). Résonance sonore et cavités supralaryngées. In P. Auzou, V. Rolland, S. Pinto & C. Ozsancak (dir.), *Les dysarthries*, pp.101-110. Marseille : Solal.

Hansen A. B. (1994). Etude du E caduc – stabilisation en cours et variations lexicales. *Journal of French and Language Studies*, 4, 25-54.

Harrington J. & S. Cassidy (1999). *Techniques in Speech Acoustics*. Dordrecht: Kluwer Academic.

Holst T. & F. Nolan (1995). The influence of syntactic structure on [s] to [ʃ] assimilation. In B. Connell & A. Arvaniti (eds), *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*, pp. 315-333.

Johnson K. (1997). *Acoustic and Auditory Phonetics*. London: Balckwell.

Johnson K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (eds), *Spontaneous Speech: Data and Analysis*, pp. 29-54. Tokyo: The National International Institute for Japanese Language.

Keating P., Cho T., Fougeron C. & C. Hsu (2003). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden & R. Temple (eds), *Papers in Laboratory Phonology VI: Phonetic Interpretation*, pp. 143-161, Cambridge: Cambridge University Press.

Kohler K. (1992). Gestural reorganization in connected speech: a functional viewpoint on 'articulatory phonology'. *Phonetica*, 49, 205-211.

Kohler K. (1995). The realization of plosives in nasal/lateral environments in spontaneous speech in German. *Proceedings of the 13th International Congress of Phonetic Sciences*, vol. 2, pp. 210-213. Stockholm.

Kohler K. (1998). The disappearance of words in connected speech. *Zas Working Papers in Linguistics*, 11, 21-34.

Ladefoged P. (1996). *Elements of Acoustic Phonetics*. Chicago: The University of Chicago Press.

Ladefoged P. & I. Maddieson (1996). *The Sounds of World Languages*. Oxford: Blackwell.

Landercy A. & R. Renard (1975). *Éléments de phonétique*. Bruxelles : Didier.

Liberman A. M. & I. G. Mattingly (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.

Liberman A. M., Delattre, P., Cooper, F. S. & L. J. Gerstman (1954). The role of consonant-vowel transition in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied*, 68, 1-13.

Lindblom B. (1963). Spectrographic study of vowel reduction. *Journal of Acoustical Society of America*, 35, 1773-17781.

Lindblom B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J.J. Jaeger (eds), *Experimental Phonology*, pp. 13-44. New–York: Academic Press.

Lindblom B. (1990). Expaining phonetic variation : a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (eds), *Speech Production and Speech Modelling*, pp. 403-439. Dordrecht: Kluwer Academic.

Longchamp F. (1988). *Etude sur la production et la perception de la parole, les indices acoustiques de la nasalité vocalique, la modification du timbre et de la fréquence fondamentale.* Thèse de doctorat d'Etat. Université de Nancy II.

Lucci V. (1983). *Etude phonétique du français contemporain à travers la variation situationnelle*. Grenoble : Editions de l'Université de Grenoble.

Machač P. & R. Skarnitzl (2009). *Principles of phonetic segmentation*. Prague: Epocha Publishing House.

MacNeilage P. F. (1998). The Frame/Content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21: 499-546.

Maeda, S. 1993. Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In M.K. Huffman & R.A. Krakow (eds), *Nasals, Nasalization and the Velum*, pp. 147-167. San Diego: Academic Press.

Marchal A. (1980). *Les sons et la parole*. Montréal : Guérin.

Marchal A. (1988). Coproduction: evidence from EPG data. *Speech Communication*, 7, 287-295.

Martin P. (2008). *Phonétique acoustique*. Paris : Armand Colin.

Mattingly I. G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, J. Anderson (eds), *The Cognitive Representation of Speech*, pp. 415–420. Amsterdam: North Holland.

Meunier C. (1994). *Les groupes de consonnes : problématique de la segmentation et variabilité acoustique*. Thèse de doctorat. Université de Provence, Aix-en-Provence.

Meunier C. & R. Espesser (2011). Vowel reduction in conversational speech in French: the role of lexical factors. *Journal of Phonetics*, 39(3):271-278.

Meynadier Y. (2001). La syllabe phonétique et phonologique : une introduction. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence*, 20, 91-148.

Meynadier, Y. (2010). *Interaction entre articulation linguopalatale et prosodie en français*. Sarrebruck: Editions Européennes Universitaires.

Moore B. (1989). *Introduction to the Psychology of Hearing*. London: Academic Press.

Nguyen N. (2001). Rôle de la coarticulation dans la reconnaissance des mots. *L'Année Psychologique*, 101, 125-154.

Nguyen N. & Z. Fagyal (2008). Acoustic aspects of vowel harmony in French. *Journal of Phonetics*, 36: 1-27.

Niebuhr O., Lancia L. & C. Meunier (2008). On place assimilation in French sibilant sequences. *Proceedings of the 8th International Seminar on Speech Production*, pp. 221-224. Strasbourg.

Ohala J.J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick & M. F. Miller (eds), *Papers from the Parasession on Language and Behavior*, pp. 178-203. Chicago: Chicago Linguistic Society.

Racine I. & F. Grosjean (2002). La production du E caduc facultatif est-elle prévisible ? Un début de réponse. *Journal of French and Language Studies*, 12, 307-326.

Recasens D. & A. Espinosa (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of Acoustical Society of America*, 125, 2288-2298.

Ridouane R., Meynadier Y. & C. Fougeron (2011). La syllabe : objet théorique et nature physique. In L.-J. Boé & J.-L. Schwartz (éd.). *La Parole : pluridisciplinarité et relations entre la substance et la forme*. *Faits de Langue*, 37 : 225-246.

Rousset I. (2004). *Structures syllabiques et lexicales des langues du monde, Données, typologies, tendances universelles et contraintes substantielles*. Thèse de Doctorat de 3e cycle, Grenoble, Université Stendhal.

Saltzman E. (1986). Task dynamic coordination of the speech articulators: a preliminary

model. *Experimental Brain Research Series*, 15, 129-144

Schwartz J.-L., Basirat, A., Ménard, L. & M. Sato (à paraître). The Perception for Action Control Theory (PACT): a perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*.

Shadle C. H. (1990). Articulatory-acoustic relationships in fricative consonants. In W. J. Hardcastle & A. Marchal (eds), *Speech Production and Speech Modelling*, pp. 187-209. Dordrecht: Kluwer Academic.

Stevens K.N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3-46.

Stevens K. N. (1998). *Acoustic Phonetics*. Cambridge: MIT Press.

Strange W. (1989). Evolving theories of vowel perception. *Journal of Acoustical* Society *of America*, 85, 2081-2087.

Straka G. (1963). La division des sons du langage entre voyelles et consonnes peut-elle être justifiée ? *Travaux de Linguistique et de Littérature de Strasbourg*, I, 17-99.

Straka G. (1964). L'évolution phonétique du latin au français sous l'effet de l'énergie et de la faiblesse articulatoires. *Travaux de Linguistique et de Littérature de Strasbourg*, II, 17-98.

Tranel B. (1987). *The Sounds of French*. Cambridge: Cambridge University Press.

Toda M. (2009). *Étude articulatoire et acoustique des fricatives sibilantes*. Thèse de doctorat. Université de la Sorbonne Nouvelle, Paris.

Tubach J-P. (1989). *La parole et son traitement automatique*. Paris : Masson.

Vallée N. (1994). *Systèmes vocaliques : de la typologie aux prédictions*. Thèse de doctorat. Université Stendhal, Grenoble.

Wright R. (2004). A review of perceptual cues and cue robustness. In B. Hayes R. Kirchner & D. Steriade (eds), *Phonetically based phonology*, pp. 34–57. Cambridge: Cambridge University Press.

Zerling J-P. (1984). Phénomènes de nasalité et de nasalisation vocaliques : étude cinéradiographique de deux locuteurs. *Travaux de l'Institut de Phonétique de Strasbourg*, 16, 241-266.

FIGURES

Figure 1.1 Oscillogrammes (en haut : détail de la forme de l'onde) et spectres d'un son périodique simple *(a-b)*, complexe *(c-d)*, apériodique continu *(d-e)* et impulsionnel *(g-h)*.

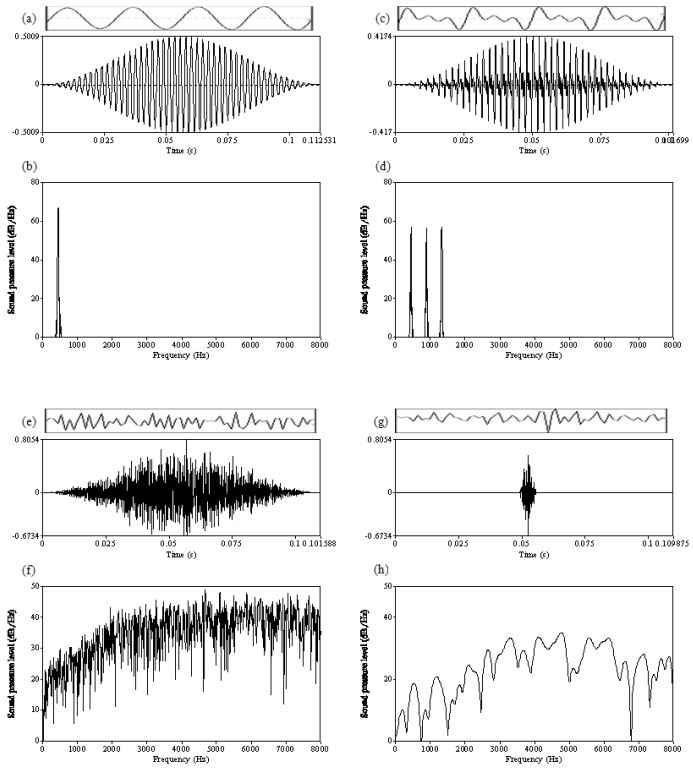


Figure 1.2 Décomposition fréquentielle (à gauche : oscillogramme, à droite : spectre) d'un son périodique complexe de 200 Hz (en gras en *a*) selon ses trois composantes périodiques simples d'amplitude différence (en pointillés en *a*) : *(b)* F_0 de 200 Hz, *(c)* 2è harmonique de 400 Hz, *(d)* 3è harmonique de 600 Hz ; adapté de Landercy & Renard (1975).

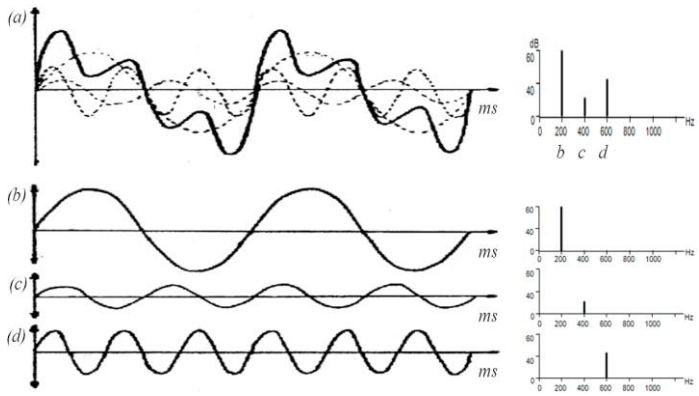
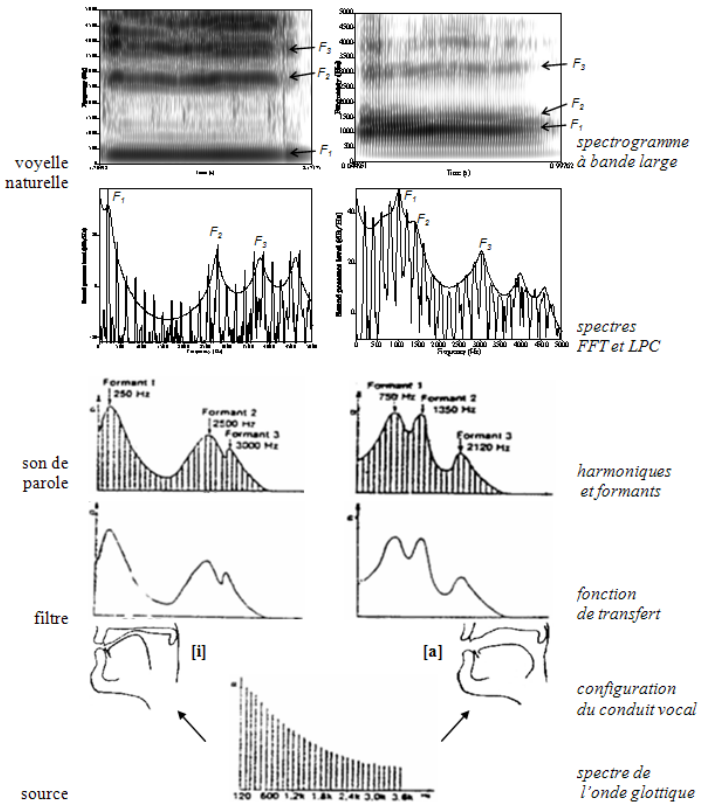
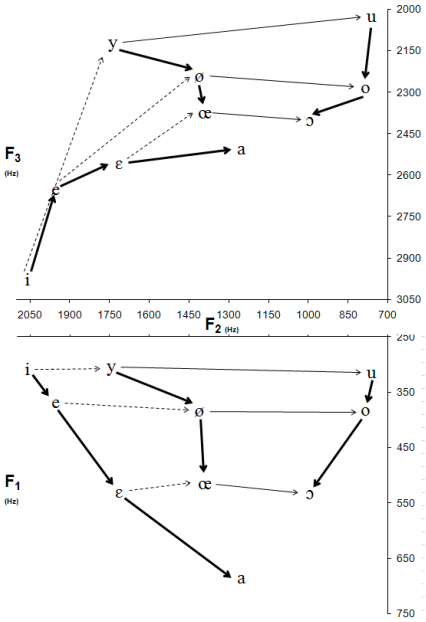


Figure 1.3 Schématisation du modèle *source-filtre* illustrant l'émergence des trois premiers formant des voyelles [i] et [a] (adapté de Landercy et Renard 1975)[26]. Au-dessus, spectres FFT et LPC superposés, et spectrogramme à bande large de ces voyelles produites par une femme.



[26] A noter dans ce schéma, que le spectre de l'onde glottique présente une pente de -12 dB par octave et que les phénomènes d'impédance du conduit et de radiation acoustique aux lèvres (de +6 dB par octave) sont intégrés à la fonction de transfert du filtre. Pour plus de détails, voir notamment Harrington & Cassidy (1999).

Figure 1.4 Triangle vocalique de l'API (en encadré) et fréquence moyenne (en Hz) du F_1, F_2 et F_3 des voyelles orales du français produites par 10 hommes en mots isolés (Tubach 1989). Les flèches épaisses indiquent l'effet de l'aperture, fines l'effet du recul du lieu d'articulation, et en pointillés l'effet de l'arrondissement.



	F_1	F_2	F_3
i	308	2064	2976
y	300	1750	2120
u	315	764	2027
e	365	1961	2644
ø	381	1417	2235
o	383	793	2283
ɛ	530	1718	2558
œ	517	1391	2379
ɔ	531	998	2399
a	684	1256	2503

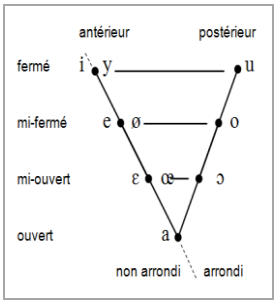


Figure 1.5 Fréquence moyenne (en Hz) du F_1 et F_2 des voyelles orales du français (i) en dialogue spontané selon leur durée *(lignes grises)* et (ii) en mots monosyllabiques isolés *(ligne noire)*. (i) valeurs issues de l'étude de Meunier & Espesser (2011) sur le CID (Bertrand et al. 2008) : 10 femmes, 46135 voyelles, E pour [e ɛ], O pour [o ɔ] et @ pour [ø œ]. (ii) valeurs issues de Tubach (1989) : 9 femmes, 180 voyelles.

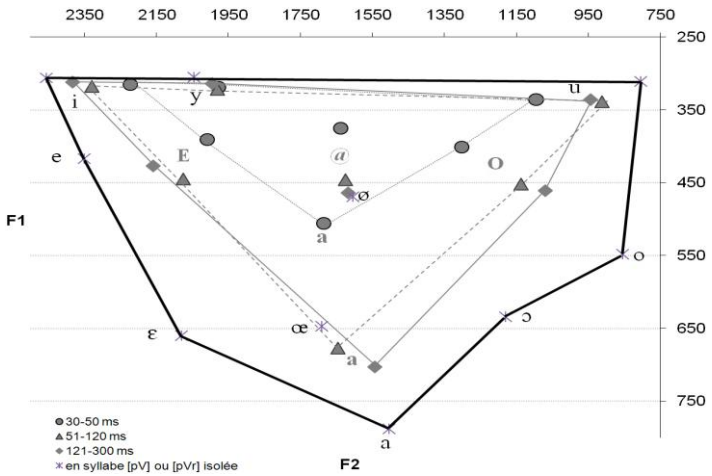
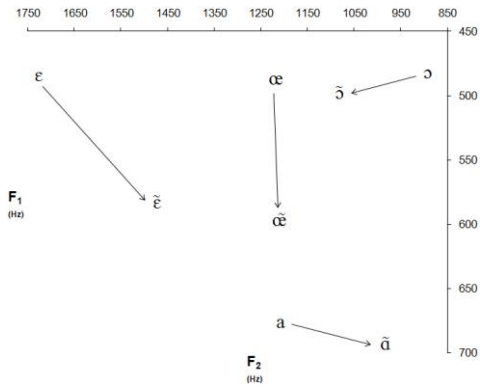


Figure 1.6 Fréquence moyenne (en Hz) du F_1, F_2 et F_3 des voyelles nasales du français : 2 hommes, 3 répétitions, logatomes (Delvaux 2002).



	F_1	F_2	F_3
ɛ	483	1728	2397
ɛ̃	583	1477	2451
œ	488	1227	2179
œ̃	596	1224	2238
ɔ	483	895	2258
ɔ̃	499	1082	2319
a	678	1200	2392
ɑ̃	692	987	2203

Figure 1.7 Oscillogramme et spectrogramme des consonnes sonantes (approximantes, vibrantes et nasales) françaises produites par un homme (lecture, logatomes, contexte [a_a]).

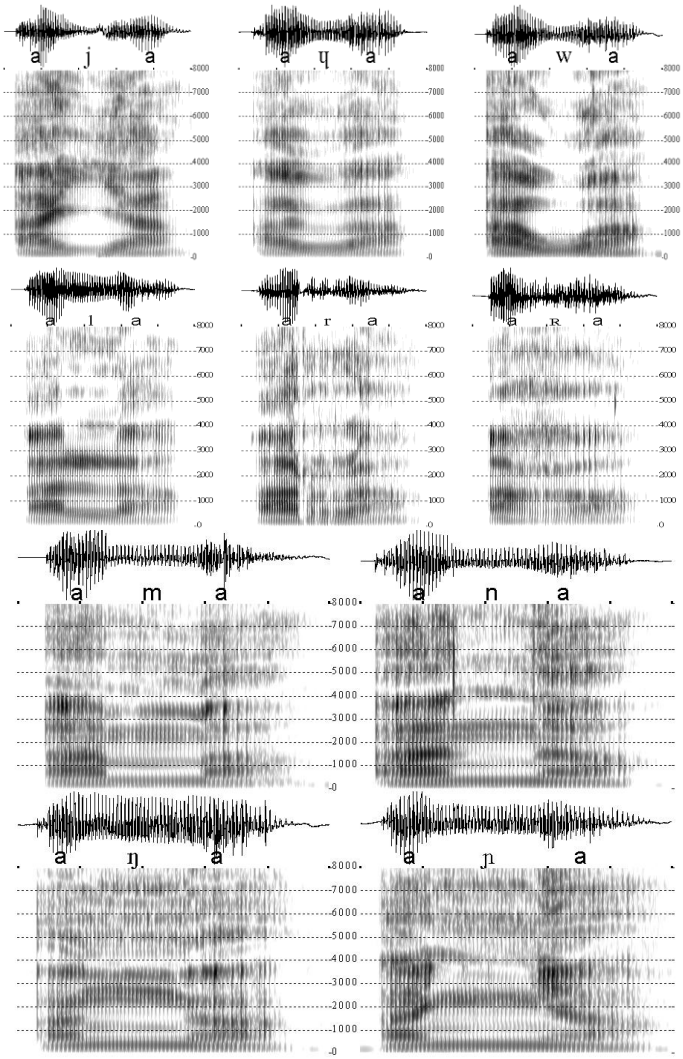


Figure 1.8 Oscillogramme et spectrogramme des consonnes obstruantes (fricatives et plosives) françaises produites par un homme (lecture, logatomes, contexte [a_a]).

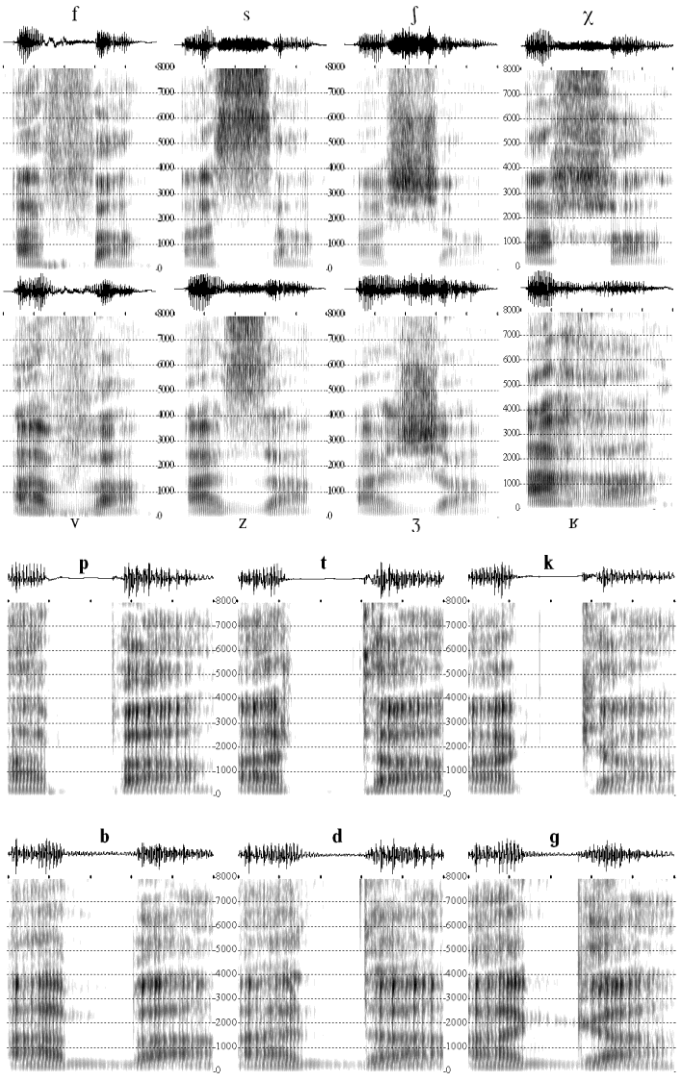


Figure 1.9 *(a)* Transitions consonne-voyelle de F_1, F_2 et F_3 *(surlignées en blanc)* pour chaque voyelle (V) et *(b)* locus de F_2 pour les lieux labial, alvéolaire et vélaire (Delattre et al. 1955). La ligne verticale symbolise la frontière entre la consonne et la voyelle. Les pointillés correspondent à l'interpolation des trajectoires pendant le bruit de friction et d'explosion de la consonne. Les valeurs formantiques des voyelles du français sont tirées de Tubach (1989).

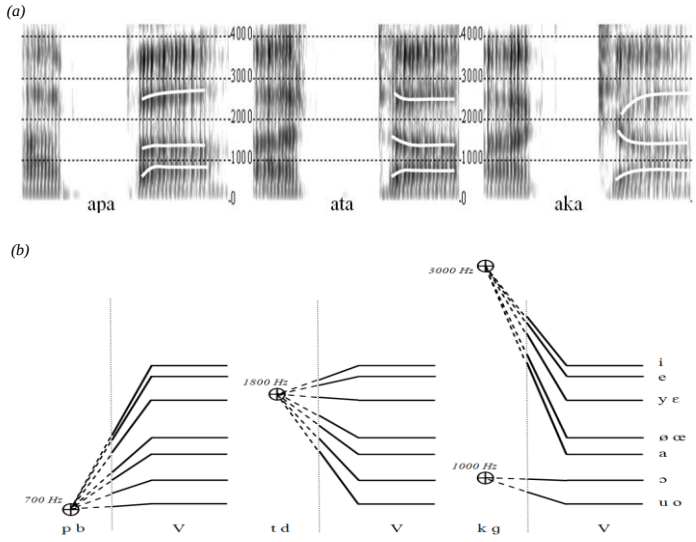


Figure 1.10 Cas de coarticulation illustrant l'influence de l'arrondissement labial de la voyelle [y] (*vs* [e]) sur la fréquence du bruit de la fricative [s] (« et ouais c'est sûr ouais », locuteur méridional, extrait du CID)

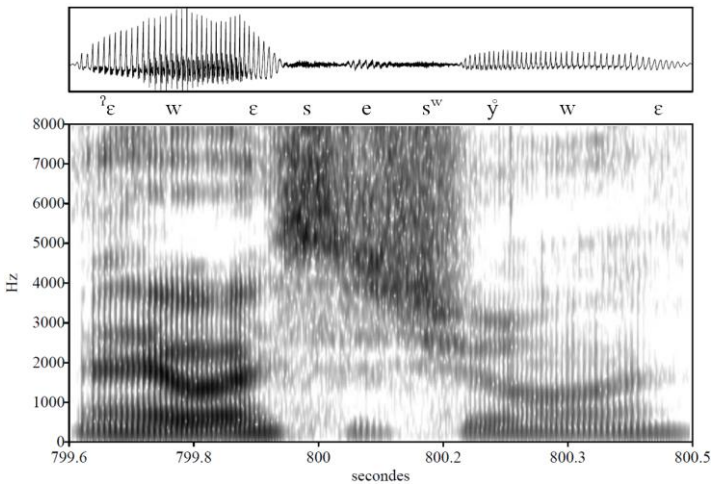


Figure 1.11 Cas de dévoisement de [l] après la plosive sourde [p] (« je sais plus », locuteur méridional, extrait du CID)

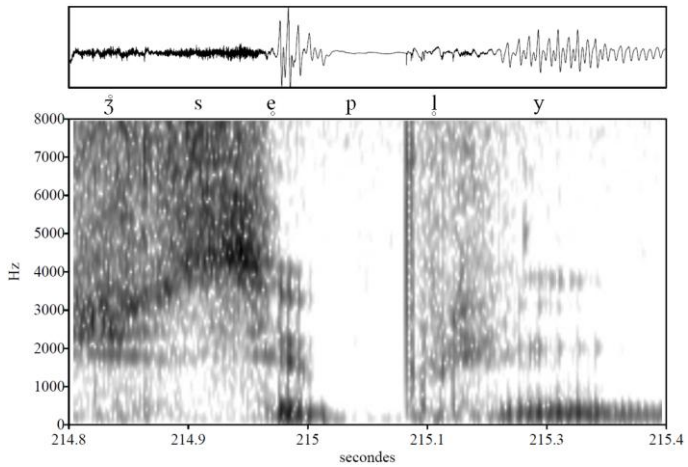


Figure 1.12 Cas de métaplasme en parole conversationnelle (« effectivement », locuteur méridional, extrait du CID)

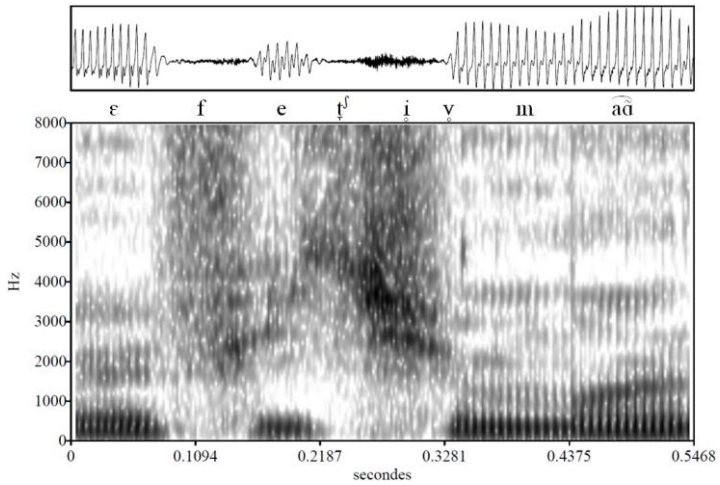


Figure 1.13 Cas de métaplasme avancé en parole conversationnelle (« tu vois », locuteur méridional, extrait du CID)