



HAL
open science

VARIATION DE LA PAROLE: CONTRAINTES LINGUISTIQUES ET MECANISMES D'ADAPTATION

Christine Meunier

► **To cite this version:**

Christine Meunier. VARIATION DE LA PAROLE: CONTRAINTES LINGUISTIQUES ET MECANISMES D'ADAPTATION. Linguistique. Université Lumière Lyon 2, 2014. tel-02427041

HAL Id: tel-02427041

<https://hal.science/tel-02427041>

Submitted on 3 Jan 2020

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

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



VARIATION DE LA PAROLE: CONTRAINTES LINGUISTIQUES ET MECANISMES D'ADAPTATION

Volume 1 – Document de synthèse

Dossier présenté pour l'obtention d'une

Habilitation à Diriger des Recherches

Par

Christine Meunier

Chargée de Recherche au C.N.R.S.

Aix-Marseille Université

Laboratoire Parole et Langage – UMR 7309

Soutenu le 19 décembre 2014 devant:

Martine ADDA-DECKER, Directeur de Recherche, CNRS, Laboratoire de Phonétique et Phonologie, UMR 7018, Paris – *Rapporteur 1*

Ann BRADLOW, Professeur, Northwestern University, Department of Linguistics, Evanston, Chicago, US – *Rapporteur 1*

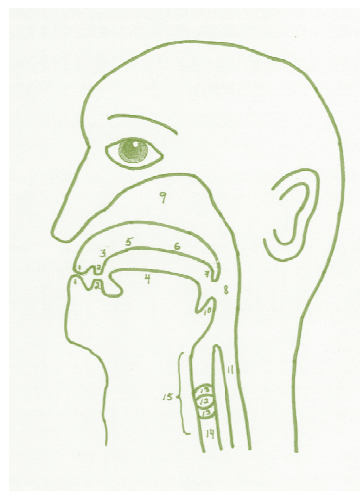
Valerie HAZAN, Professeur, University College London, Faculty of Brain Sciences, Division of Psychology and Language Sciences, Department of Speech, Hearing and Phonetic Sciences, London, UK – *Rapporteur 2*

Noël NGUYEN, Professeur, Aix-Marseille Université, Laboratoire Parole et Langage, UMR 7309, Aix-en-Provence – *Examineur*

François PELLEGRINO, Directeur de Recherche, CNRS, Laboratoire Dynamique du Langage, UMR 5596, Lyon – *Tuteur*

"A celui à qui je dois tout"
Claire Brétécher, *Les Frustrés*

Remerciements



© 2002 Emily J. Fajardo

Rares sont les occasions de remercier ceux qui ont marqué notre parcours, ceux qui nous accompagnent et nous entourent quotidiennement ou à l'occasion. Les remerciements d'un mémoire peuvent paraître assez convenus ou finalement assez conventionnels. Conventionnels d'un côté, mais aussi très personnels. Ils nous permettent, à un moment donné, de nous arrêter pour regarder autour de nous, regarder les autres et réaliser que leur présence est ce qui a permis ce que nous faisons et ce que nous sommes. Même si, la plupart du temps, ils n'en n'ont pas conscience...

Certaines personnalités, par leur charisme, leurs qualités humaines et professionnelles ont façonné mon parcours et mon regard sur la recherche. Je veux parler d'Uli Frauenfelder et Alain Content qui m'ont accueillie à Genève pour mon post-doctorat, une merveilleuse expérience humaine et professionnelle. Mais également, les trois directeurs successifs du Laboratoire Parole et Langage. Tout d'abord Bernard Teston qui a eu confiance en moi (bien plus que moi!). Philippe Blache, bien sûr. Je n'ai jamais connu quelqu'un d'aussi entraînant que lui. Philippe est une fusée que l'on a envie de suivre, coûte que coûte. Une fusée dont, hélas, rares sont ceux qui perçoivent la sensibilité. Et puis Noël Nguyen. Nos chemins sont parallèles depuis tellement d'années. La direction du LPL nous a rapprochés et je ne pouvais imaginer une meilleure configuration en tant que directrice adjointe. Certaines me comprendront sans doute, il n'est pas toujours facile d'être une femme et de se sentir légitime dans des postes de responsabilité. Noël fait passer tous ces doutes. Alors merci pour cela, c'est énorme.

Je voudrais aussi remercier les membres du jury de cette HDR. Je suis extrêmement heureuse de la composition de ce jury qui est pour moi parfaitement représentatif de mes centres d'intérêt. Merci à Ann Bradlow et à Martine Adda-Decker qui ont accepté de lire ce document et d'en être les rapporteuses. Merci à Valerie Hazan et aussi à Ann d'avoir accepté de faire le trajet depuis l'étranger. Merci bien sûr à Noël de m'accompagner de nouveau dans cette étape. Et surtout, merci à François Pellegrino pour son

extrême disponibilité et son efficacité. Tout est simple et rapide avec François. Merci pour ta gentillesse, tes encouragements et la pertinence de tes remarques.

Quelques mots ici pour la *Communauté Parole francophone*, et particulièrement l'AFCP. Issue de milieux disciplinaires variés cette communauté a su construire une convivialité dans laquelle il fait vraiment bon évoluer. Profitons de cette richesse.

Je me suis souvent dit que le luxe absolu, plus que tout autre critère, était de se sentir bien dans son milieu de travail. Je pense souvent à ceux qui se rendent chaque jour à leur travail en redoutant les personnes qu'ils vont y croiser. Le LPL n'est évidemment pas le paradis. Comme tout environnement de travail, il y a parfois des tensions. Pourtant quelle chance, quel luxe de venir travailler pour y retrouver des amis, de vrais amis. Je pense bien sûr à Pascal, Philippe, Noël, Alain, Marion, Béa, Yohann, Roxane, etc. et à leurs moitiés, Marianne, Gino, Mali, Mylène. Et puis à Robert, bien sûr, à mi-chemin entre Ronchin et Mission Croco. Robert nous rappelle à tous que l'humour n'est pas une chose légère. C'est juste une question de survie...

Merci aussi à l'ensemble des personnels administratifs et techniques de notre laboratoire (administration, technique, documentation, etc.) sans lesquels nous ne pourrions pas évoluer correctement.

Il y a aussi celles qui sont à *l'extérieur*, mais si près. Quel luxe, là encore, de travailler avec des amies, de faire se mélanger le travail, le bien être et l'humour! Je parle des Corinnes (Astésano et Fredouille). Et bien sûr de Cécile, *ma copine de toujours* avec qui j'espère encore partager longtemps l'amitié et (éventuellement) le travail... ;-)

Une petite (grosse!) place pour mon Yoh qui a accepté, à ses risques et périls, de partager avec moi le nid et le couvert. Je n'ai jamais vu un ours aussi attentionné... Et merci aussi à Marine. Merci à vous, je crois que j'ai de la chance...

Lorsque l'on a un travail qui prend de la place il est judicieux (et même très raisonnable) d'avoir un *contrepoids*. Des personnes extérieures, qui vous sortent de là et vous permettent de relativiser un peu... En ce qui me concerne, j'ai une "tribu" composée de Mumu, Sylvie, Dom, Eric, Caro, Feuille, Linecé, Martine et Valérie... Ma tribu me permet de respirer, de respirer la vie, de manger la vie, d'aimer le genre humain... Un équilibre indispensable. Je les en remercie.

Mais celui qui m'aide le plus à relativiser est probablement Samuel, mon fils. Je mesure à quel point la clairvoyance habite cet enfant lorsqu'il m'apostrophe ainsi: "*Mais c'est nul quand même ce que tu fais. Ça sert à rien, en fait*"... A ces mots, mon orgueil est à son paroxysme tant je suis fier d'avoir inculqué à mon fils le refus de l'hypocrisie (un peu trop, peut-être...), le sens critique et la désacralisation du savoir... Merci Samuel pour ton honnêteté. Garde-la. Merci aussi pour ta patience. Et surtout, merci d'être là. ...

Et puis mes pensées iront pour finir à Yza, qui a appris à me supporter et à m'accompagner (trop peu) depuis quelques années maintenant. Mais tout cela va changer...

TABLE DES MATIERES

SYNTHESE DES ACTIVITES DE RECHERCHES	4
I. Introduction	4
I.1 Un chemin dans la parole.....	4
I.2 Structure du document.....	9
II. Approches méthodologiques	11
II.1 Les corpus et leur exploitation.....	12
II.1.1 L'annotation des corpus oraux.....	14
II.2 Approche psycholinguistique: la reconnaissance des mots.....	30
II.3 Bilan.....	35
III. Parler: les sons de la parole	37
III.1 Les sons et leur usage.....	39
III.2 Variations et réductions en parole.....	46
III.2.1 Coarticulation et assimilation.....	48
III.2.2 Phénomènes de réductions dans la parole.....	55
III.3 Bilan.....	64
IV. S'adapter	67
IV.1 Aux langues.....	69
IV.1.1 Systèmes phonologiques et variations phonétiques.....	69
IV.1.2 Le lien perception-production dans l'apprentissage d'une langue seconde.....	73
IV.2 Aux situations de parole.....	75
IV.2.1 Information et variation de la parole.....	78
IV.2.2 La réduction vocalique.....	80
IV.3 Aux contraintes.....	94
IV.3.1 Les populations pathologiques.....	94
IV.3.2 La parole dysarthrique.....	96
IV.4 Bilan.....	102
V. Perspectives: relations entre propriétés des systèmes linguistiques et mécanismes d'adaptation	103
V.1 Les sons des langues et leur usage.....	103
V.2 Interprétation des formes de réduction.....	105
V.3 Mécanismes d'adaptation: confrontation des systèmes.....	106
V.4 Perception en parole spontanée.....	107
VI. Références	109

SYNTHESE DES ACTIVITES DE RECHERCHES

Puisque ces mystères me dépassent, feignons d'en être l'organisateur.

Jean Cocteau, *Les Mariés de la Tour Eiffel*

I. INTRODUCTION

Ce document regroupe, comme il est convenu, une synthèse de mes travaux de recherche depuis mon doctorat, autrement dit depuis vingt ans. J'ai fait le choix d'en faire un regroupement de toutes les réflexions et questionnements qui se sont présentés à moi au cours de ces années, au gré des rencontres et des projets développés.

J'ai très longuement hésité à faire apparaître le mot **Variation** dans le titre. J'éprouve un certain malaise à utiliser ce terme tant il est imprécis, galvaudé, mais aussi totalement incontournable. Un contenu très imprécis enrobe ce mot et le rend ainsi très peu explicite. Comment parler de parole, ou même de sciences humaines, sans aborder la notion de *variation*? La variation est partout et, de ce fait, elle n'est plus nulle part. Elle ne reflète plus qu'un système en mouvement. Un système dans lequel l'adaptation devient centrale.

I.1 UN CHEMIN DANS LA PAROLE

Il est un moment clef dans mon parcours universitaire qui détermine le point de départ de mon intérêt pour la parole. D'une certaine façon, il s'agit d'un point méthodologique. De formation plutôt littéraire, j'assistais à un cours sur la langue espagnole. Sous forme de diversion, un de nos professeurs nous a alors appris que des chercheurs travaillaient sur les *sons des langues* et qu'ils mesuraient *physiquement* leurs caractéristiques. Cette possibilité de relier les langues et des mesures physiques objectives m'a alors totalement fascinée. C'était la possibilité d'associer les sciences humaines et les sciences "dures" et surtout, de ne pas observer qu'une seule de ces deux facettes.

Cette fascination ne m'a pas quittée. Le langage est, à ce titre, un objet particulièrement captivant relevant à la fois de propriétés totalement abstraites et immatérielles, de processus cognitifs complexes, mais aussi composantes physiologiques observables, mesurables,

quantifiables. Ces différentes facettes sont tour à tour défendues ou vilipendées en fonction de nos orientations disciplinaires. En ce qui me concerne, c'est cette ambivalence, cette complémentarité, cette interaction qui font tout l'intérêt de cet objet qu'est le langage. Plus encore, cet objet est partagé par l'ensemble de l'humanité. Chaque individu manipule un objet partagé par tous et pourtant, il le manipule avec singularité. Chaque langue est spécifique et comporte des dialectes à l'intérieur desquels chaque communauté comporte ses propres codes que chaque individu manipule avec ses propres caractéristiques. Et pourtant le langage a cette universalité qui fait qu'avec un peu d'adaptation, nous sommes capables de parler avec n'importe quel individu de la planète.

La vision que j'ai du langage est, comme tout un chacun, conditionnée par mes expériences professionnelles et mes rencontres au sein des différents laboratoires ou projets. A une formation doctorale très concentrée sur la description phonétique et sur l'analyse des formes de variation phonétiques a succédé un post-doctorat à l'université de Genève au sein de l'équipe d'U.H. Frauenfelder en 1994 dans le cadre d'un projet financé par le Fond National de la Recherche Scientifique (Suisse) "*La perception du langage parlé: la reconnaissance des mots*". L'**approche psycholinguistique** de ce laboratoire m'était alors tout à fait étrangère. J'ai donc énormément appris à cette époque sur les processus de traitement perceptif de la parole et sur les modèles de reconnaissance des mots parlés. C'est aussi à de toutes nouvelles méthodologies que j'ai été confrontée. Leur apprentissage m'a ouvert de nombreuses perspectives. L'environnement à la Faculté de Psychologie et des Sciences de l'Education (Université de Genève) était tout à fait propice pour des échanges avec d'autres composantes de la psychologie et j'ai beaucoup élargi mon champ de connaissance durant ces quatre années. Plus précisément, la collaboration avec U.H. Frauenfelder et A. Content fut extrêmement enrichissante et formatrice pour moi aussi bien professionnellement qu'humainement.

Suite à mon recrutement au CNRS, j'ai eu la chance, en 2000, d'obtenir un financement dans le cadre du programme Cognitique (Ministère de la Recherche) pour le projet "*Densité des systèmes vocaliques et traitement cognitif des unités sonores dans différentes langues*". C'était la première fois que je me confrontais à une **comparaison inter-langue** et les résultats obtenus dans ce projet ont stimulé mon intérêt et mon questionnement concernant le lien entre les phénomènes de variation et la spécificité des systèmes linguistiques. A la lumière de travaux plus récents (pathologie du langage), les travaux multilingues ont pour moi un nouvel attrait.

En 2007, j'ai eu la chance de participer au projet Européen *Sound to Sense*¹ (Marie Curie Research Training Network, dirigé par Sarah Hawkins, Cambridge, UK). Ce projet regroupait 14 laboratoires européens spécialisés en traitement automatique de la parole, psychologie, phonétique, apprentissage L2. L'objectif général du projet était de proposer des modèles de traitement de la parole (reconnaissance automatique de la parole, synthèse automatique de la parole) qui reflétaient précisément **la forte flexibilité et la robustesse** du traitement de la parole par les sujets humains. Il visait, dans le même temps, l'élaboration de nouveaux modèles théoriques pour l'acquisition des langues secondes ainsi que pour le diagnostique et le traitement des pathologies de la parole. Ce projet très ambitieux avait pour fonction de former de jeunes chercheurs en leur permettant de se déplacer dans le réseau autour d'un projet scientifique décliné sous plusieurs axes. A cette occasion, j'ai pu encadrer Oliver Niebuhr (actuellement Junior Professor à l'Université de Kiel en Allemagne), dans le cadre du projet *Perceptual salience of Fine Phonetic Details: Cross-linguistic comparison of contextual sensitivity*. Cette collaboration fut très fructueuse. C'est aussi dans le cadre de ce projet que j'ai pu croiser des chercheurs comme Sven Mattys (University of Bristol, UK), Mirijam Ernestus (MPI, Nijmegen, Pays-Bas) ou Richard Ogden (York, UK) dont les différentes approches ont considérablement élargi mon champ d'investigation.

De 2007 à 2010, j'ai eu l'opportunité d'assurer, avec Sibylle Kriegel, la co-responsabilité de l'équipe "Dynamique des systèmes linguistiques, créoles et variétés émergentes". Elle regroupait des créolistes, des sociolinguistiques, des spécialistes de l'apprentissage des langues et des analystes du discours textuels. Cette équipe, assez hétérogène, m'a fait côtoyer les problématiques liées au plurilinguisme ou aux facteurs socio-dialectaux et impliquées dans le **changement linguistique**. Ces questions, assez éloignées de ma propre formation, furent très enrichissantes et ont suscité en moi l'intérêt pour la prise en compte des variations indexicales ainsi que de la situation linguistique propre au locuteur.

De 2009 à 2012, j'ai eu la chance de participer à deux projets financés par l'ANR (Agence National de la Recherche). Le premier, OTIM (*Outils de traitement d'information multimodale*, PI: P. Blache) avait pour ambition de développer des outils permettant un requêtage des informations linguistiques sur plusieurs niveaux (du niveau phonétique au niveau gestuel, en passant par la syntaxe, la prosodie, et discours, etc.). Ce projet de **mise en relation des différents niveaux linguistiques** m'a permis d'entrevoir comment il était possible de relier des

¹ <http://www.sound2sense.eu/>

informations phonétiques fines à des unités linguistiques de plus haut niveau et couvrant des domaines très larges. C'était aussi pour moi l'occasion de travailler sur un type de parole très relâché (CID, corpus de parole conversationnelle). Le second projet DesPhoAPaDy (*Description Phonetico-Acoustique de la Parole Dysarthrique*, PI: C. Fougeron) visait à structurer une base de données de **parole dysarthrique** en vue d'une description acoustico-phonétique différenciée des pathologies associées à la dysarthrie. Ce fut pour moi un premier contact avec la parole pathologique et avec la complexité des facteurs de variation à maîtriser dans ce cadre de recherche. Ces deux projets n'avaient, au premier abord, pas de lien, mais j'ai pu faire le parallèle entre l'analyse des phénomènes de réduction de la parole spontanée (par le biais de l'annotation phonétique dans le cadre d'OTIM), d'une part, et les déviances phonétiques de la parole pathologique, d'autre part. La similarité de ces phénomènes m'a poussé à développer un projet dans lequel ces différents aspects étaient mis en perspective: les populations pathologiques et non-pathologiques; la variation dans la production de la parole dans des styles variés. C'est dans ces conditions que le projet TYPALOC (*Variations normales et anormales de la parole: Typologie, Adaptation, Localisation*, PI: Christine Meunier), financé par l'ANR, a commencé en novembre 2012.

J'ai eu l'occasion de m'intéresser à la **parole pathologique** alors que je donnais des cours à l'école d'orthophonie de la faculté de médecine à Marseille dès 2001. Par la suite, j'ai participé aux enseignements de Master "Pathologie du langage" à l'Université de Provence. Ces divers enseignements m'ont amenée à m'interroger sur les indices phonétiques qui permettent de délimiter les contours de la parole normale de ceux de la parole pathologique, malgré la très grande variabilité phonétique observée en parole normale. En 2010, lors du changement de contrat du laboratoire, j'ai intégré l'équipe "Parole : Contraintes, Variations et Structures" animée par P. Welby et A. Giovanni. Cette équipe est composée de phonéticiens, cliniciens de chercheurs en neurosciences. Les discussions régulières avec les cliniciens de cette équipe (A. Giovanni, D. Robert, J. Revis, A. Lagier, etc.) ainsi que les collaborations que nous entretenons dans le cadre des projets de recherche ou de l'encadrement des mémoires d'orthophonie sont extrêmement fructueux et particulièrement propices au développement des travaux au carrefour de la phonétique et des recherches cliniques.

Depuis 2012, le laboratoire Parole et Langage fait partie du **LABEX BLRI** (Brain and Language Research Institute¹) dont P. Blache est le responsable. Ce LABEX regroupe sept laboratoires spécialisés en psychologie, neurosciences, informatique, linguistiques, médecine. Cet environnement, là encore très propice aux échanges scientifiques, nous a permis d'obtenir une bourse de thèse en co-encadrement avec Corinne Fredouille (Laboratoire d'Informatique d'Avignon) et de déposer un projet de recherche dans le cadre d'une collaboration internationale avec Hong Kong.

En effet, depuis plus d'un an, je collabore avec Yao Yao (Polytechnic University, Hong Kong) sur un projet de recherche financé par le LABEX BLRI "*Vowel variation in conversational speech in French and English: Effects of phonological and grammatical properties of lexicon*". Ce projet est né de deux articles dans lesquels nous étions toutes les deux impliquées (Gahl et al., 2012; Meunier and Espesser, 2011). Dans ces travaux, il était question d'évaluer le rôle de certains facteurs lexicaux sur la réduction vocalique dans des corpus de parole conversationnelle. L'un se basait sur un corpus de l'anglais américain (Buckeye Corpus), l'autre sur un corpus de français (CID). Nous avons ainsi pu commencer à croiser les questions relatives à ces travaux sur les deux langues. Cette collaboration prend suite désormais dans le cadre du projet International de l'IDEX AMIDEX (Aix Marseille Université) **VARIAMU *Variations in Action: a MULTILINGUAL approach*** (dirigé par L. Prévot, LPL), en collaboration avec les universités de Hong-Kong et de Macau (Chine), auquel Y. Yao et moi-même participons. Les premiers contacts avec ces deux universités sont très enrichissants et ouvrent de nouvelles perspectives sur la possibilité de travailler sur des corpus de chinois mandarin en parole spontanée.

En tout dernier lieu, j'évoquerai mon rôle de directrice adjointe du Laboratoire Parole et Langage depuis 2011. Cette fonction est bien sûr administrative mais l'organisation d'une structure comme le LPL permet une vision relativement large de la possible complémentarité des recherches sur le langage. Le LPL, originellement *Institut de Phonétique*, s'est considérablement élargi depuis une quinzaine d'années, aussi bien en nombre de personnel qu'en diversité thématique. Il accueille désormais, outre des phonéticiens et des linguistes, des psycholinguistes, des informaticiens, des cliniciens, des gestualistes, des sociolinguistes, des neuro-scientistes, etc. La rencontre de ces disciplines n'est pas toujours aisée. Mais lorsqu'elle est possible, elle donne lieu à des recherches d'une grande originalité et à de très riches discussions.

¹ <http://www.blri.fr/>

L'ensemble de ces rencontres et de ces expériences a façonné mes orientations de recherches, mes questionnements et mes hypothèses. Le document développé ci-dessous en est un recueil organisé en fonction des points qui me semblent fondamentaux pour l'ensemble de mon travail.

I.2 STRUCTURE DU DOCUMENT

Le document ci-dessous est structuré en quatre parties: 1/ Approches méthodologiques; 2/ Parler: les sons de la parole; 3/ S'adapter; 4/ Perspectives. Les trois premières parties regroupent une synthèse de mes travaux dans ces différents axes. Dans la quatrième partie, j'expose mes perspectives de recherche en fonction des bilans qui ont été effectués sur les trois parties précédentes.

Le choix des trois premières parties n'est pas relatif à un parcours chronologique de mes travaux. Il correspond plutôt à des axes transversaux qui sont presque toujours présents dans mes différents projets. A ce titre, des travaux spécifiques se voient répartis dans plusieurs axes selon leur pertinence concernant l'approche méthodologique ou les notions d'unités phonétiques ou d'adaptation.

Dans la première partie, j'ai souhaité mentionner l'ensemble de mes contributions concernant les corpus oraux et leurs annotations. Si cette partie ne fait pas forcément référence à de nombreuses publications, elle est pourtant essentielle dans mon parcours professionnel. La partie méthodologique comporte également les travaux effectués à Genève et l'approche psycholinguistique de la parole. Ces travaux auraient pu être développés dans la deuxième partie. Toutefois, il m'a semblé pertinent d'articuler une réflexion autour de la complémentarité des approches de type Corpus (production) et Traitement (perception).

Les deux parties suivantes concernent, d'une façon ou d'une autre, la variation de la parole. Toutefois, j'ai voulu distinguer une partie *Description des sons de la langue*, d'une partie *Facteurs explicatifs*. Aussi, dans la partie 2 sont développées des observations concernant les sons du langage. Cette partie est volontairement descriptive et rend compte aussi bien de la distribution des phonèmes dans la langue que de leurs réalisations effectives dans la parole. La partie 3, en revanche, se veut plus interprétative. Elle propose de regarder différentes formes de variation à la lumière du mécanisme d'adaptation. Ce mécanisme se décline arbitrairement sous trois points: les langues, le type de parole et les contraintes physiologiques. Cette distinction arbitraire permet d'appréhender l'adaptation sous des angles différents. Toutefois, nous l'envisageons comme un mécanisme global de fonctionnement de la parole.

Je n'ai pas fait apparaître dans ce document des travaux dans lesquels j'estime que mon investissement a été limité. En revanche, sont présents, tous les travaux menés en collaboration dans la mesure où ils font pleinement partie des mes préoccupations scientifiques.

Chaque partie est organisée de la façon suivante: à une introduction, plus ou moins brève, font suite des synthèses relatives à mes contributions dans la thématique introduite. Ces contributions sont signalées par un encadrement grisé dans lequel sont mentionnées les publications concernées. Chacune des trois parties fait ensuite l'objet d'un bref bilan des contributions dans lequel sont mentionnés, le cas échéant, les points qui seront développés dans les perspectives.

II. APPROCHES METHODOLOGIQUES

J'ai choisi délibérément de commencer cette synthèse de mes travaux par des aspects méthodologiques. Les deux approches que je mentionnerai ici concernent les corpus et leur exploitation, d'une part, et l'approche psycholinguistique, d'autre part. Ces deux approches font partie intégrante de ma formation professionnelle, elles sont complémentaires et il se trouve que leur confrontation est au cœur des préoccupations actuelles concernant les recherches sur la parole.

Concernant les corpus, j'aborderai la question des grands corpus, notamment les corpus de parole naturelle, spontanée, et de leur exploitation. Depuis une vingtaine d'années, les linguistes tentent d'exploiter des données issues de la langue dans son usage le plus courant. Cette exigence a donné naissance à des corpus faisant émerger des phénomènes linguistiques inconnus et complexes. Comprendre ces phénomènes est un véritable défi pour les recherches actuelles sur la parole et demandent une interaction forte des différentes composantes de la linguistique. Mais il s'agit là de production de la parole.

Comprendre comment fonctionne le système perceptif relève d'approches méthodologiques différentes. Nous évoquerons ici, nos propres travaux au sein de l'équipe d'U.H. Frauenfelder à Genève portant sur la reconnaissance des mots et les unités d'accès au lexique. Ce type de méthodologie, nous le verrons, s'appuie sur des stimuli de parole très contrôlés (aussi bien dans leur choix que dans leur enregistrement) et les tâches pour lesquelles les auditeurs sont sollicités sont, elles mêmes, très loin des conditions habituelles de perception de la parole.

Ces deux approches semblent très éloignées, l'une cherchant à comprendre la langue dans son usage, l'autre, les mécanismes perceptifs recueillis sur une parole décontextualisée censés représenter les processus fondamentaux. Toutefois, des travaux récents semblent montrer qu'elles ne sont pas inconciliables. Certains projets ont ainsi pour ambition de faire le lien entre des approches expérimentales et l'étude de la langue dans ses usages. Ce fut le cas du projet européen *Sound to Sense*¹ (dirigé par Sarah Hawkins entre 2007 et 2011) dans lequel des projets de recherche tels que "Multi-linguistic and comparative research on fine phonetic detail", "Imperfect knowledge/signal", "Beyond short units of speech", et "Exemplars and abstraction"

¹ <http://www.sound2sense.eu/>

était réunis. La confrontation de chercheurs représentatifs de chacun de ces domaines a donc permis d'établir un lien entre des approches très différentes. Les travaux de Mirjam Ernestus et collègues (Ernestus, 2013, 2014; Hanique et al., 2013) sont, à ce titre, très représentatifs du lien possible entre les travaux sur la parole spontanée et l'approche expérimentale et psycholinguistique. Dans un numéro spécial de *Journal of Phonetics*¹ consacré aux phénomènes de réduction de la parole Ernestus and Warner (2011) font le constat suivant: "*Corpus studies cannot answer all questions about the production of reduced pronunciation variants, and they provide little information about speech comprehension. Research on reduction therefore also requires experimental approaches*". Enfin, plus récemment, une question habituellement dédiée aux sujets et aux méthodes issues de la psycholinguistique (l'effet du voisinage phonologique) a été traitée par le biais d'une analyse d'un corpus de parole conversationnelle, le Buckeye Corpus (Gahl et al., 2012).

C'est donc dans cette optique que j'expose mes contributions concernant ces deux approches, en faisant mien ce défi que le Laboratoire Parole et Langage s'est donné comme perspective: concilier l'approche expérimentale et l'étude de la langue dans son usage.

II.1 LES CORPUS ET LEUR EXPLOITATION

La question des corpus et des outils liés à leur exploitation sont à l'origine d'un changement fondamental concernant à la fois les approches méthodologiques mais aussi la nature des études sur la parole. En effet, depuis une vingtaine d'années, la nature des corpus exploités dans le cadre des recherches sur la parole a considérablement changé. Les phonéticiens avaient pour habitude de constituer eux-mêmes leurs corpus, en général de taille limitée, et permettant de répondre spécifiquement aux questions posées dans leurs études. Si cela est toujours le cas, l'arrivée de très grands corpus a clairement modifié la donne. Il faut probablement attribuer l'arrivée des Très Grands Corpus (TGC) aux recherches sur le traitement automatique des langues et la nécessité d'évaluer des outils sur un grand nombre de données de parole (Galliano et al., 2005). Cette essor a permis 1/ un développement considérable de la taille des corpus; 2/ un développement conjoint des outils automatiques permettant à la fois l'annotation de ces corpus mais aussi leur analyse.

¹ Ernestus, M., & Warner, N. (Eds.). (2011). Speech reduction [Special Issue]. *Journal of Phonetics*, 39(S1).

En parallèle, les linguistes ont cherché à constituer des corpus de parole plus proches des productions courantes des locuteurs. Le projet *Phonologie du Français Contemporain* (Durand et al., 2003) avait pour but de décrire l'ensemble des variantes du français produites par des francophones à l'heure actuelle. Les enregistrements comprenaient donc différents points d'enquête, mais aussi des situations de parole différentes (lecture de mots, de phrases, entretiens, etc.). L'ensemble de ces enregistrements représente donc une très grande quantité de données de parole.

Sur la même période aux Etats-Unis, le Buckeye Corpus (Pitt et al., 2005) a été constitué. Des locuteurs américains ont été enregistrés au cours de conversations (entre 30 et 60 minutes pour chaque locuteur) dans le but de recueillir des enregistrements de parole spontanée permettant de pouvoir étudier les variations phonologiques en anglais. Ce corpus représente 307 000 mots produits par 40 locuteurs de l'Ohio. Plus récemment, un corpus de dialogue conversationnel de français a été enregistré à Nijmegen (Torreira et al., 2010) de façon à étudier les caractéristiques lexicales et phonologiques de la parole conversationnelle.

La constitution de ces corpus a permis aux linguistes de focaliser leurs recherches sur des productions souvent moins contrôlées mais plus riches et plus naturelles. La quantité de données fournie par ces TGC est nécessaire pour interpréter les phénomènes observés. En effet, les productions n'étant plus contrôlées, il va de soi que les sources de variation ne le sont plus non plus. Or, dans ces situations de parole, les sources de variations sont très nombreuses (prosodie, structure du discours, interaction, etc.). Ainsi, pour décrire correctement les données phonétiques, il est indispensable de disposer d'un très grand nombre d'échantillons.

Dans la même optique, Le Corpus of Interactional Data (Bertrand et al., 2008) a été créé au Laboratoire Parole et Langage (Aix-en-Provence). Avec un nombre de locuteurs plus restreint (16), ce corpus avait pour ambition de mettre les locuteurs en situation d'interaction autour de thèmes de discussion (récits de conflits professionnels ou d'histoires insolites). Il s'agissait de conversations entre deux locuteurs se connaissant bien (donc sans intervention d'un interviewer ou d'un expérimentateur au cours de l'enregistrement). Le CID représente huit dialogues d'une heure chacun dans un processus de conversation très familière. La plupart des travaux mentionnés dans ce document prennent appui sur ce corpus très riche concernant la variation et la réduction phonétique.

II.1.1 L'annotation des corpus oraux

Les corpus évoqués ci-dessus sont inutilisables pour les linguistes s'ils ne sont pas annotés. La plupart du temps, les corpus sont transcrits orthographiquement puis alignés automatiquement en phonèmes. Tandis que pour certains corpus, cet alignement est entièrement corrigé manuellement par des experts (Pitt et al., 2005), pour d'autres la correction est partielle (Bertrand et al., 2008), ou absente (Galliano et al., 2005). Dans les contributions mentionnées ci-dessous j'évoque les différents aspects de la mise en place de l'annotation pour les corpus de parole.

Contribution:	Les annotations multi-niveaux
Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier , C., Priego-Valverde, B., Rauzy, S. (2008) "Le Cid - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle", <i>Traitement Automatique des Langues</i> , 49, 105-134.	
Blache, P.; Bertrand, R.; Bigi, B.; Bruno, E.; Cela, E.; Espesser, R.; Ferré, G.; Guardiola, M.; Hirst, D.; Magro, E.-P.; Martin, J.-C.; Meunier , C.; Morel, M.-A.; Murisasco, E.; Nesterenko, I.; Nocera P., Pallaud B., Prévot L., Priego-Valverde B., Seinturier J., Tan N., Tellier M., Rauzy S. (2010). <i>Multimodal Annotation of Conversational Data. Proceedings of Linguistic Annotation Workshop (2010 juillet 15-16 : Uppsala, SWEDEN)</i> . 2010, 6 pages.	
P. Blache, R. Bertrand, M. Guardiola, M.-L. Guénot, C. Meunier , I. Nesterenko, B. Pallaud, L. Prévot, B. Priego-Valverde, S. Rauzy (2010) "A formal annotation model as a preliminary step before annotation scheme: an experiment", <i>Proceedings of Language Resource and Evaluation Conference</i> , Malte, mai 2010.	

Mes travaux concernant l'annotation des corpus oraux se sont, en grande partie, articulés autour des travaux menés dans le cadre du projet OTIM¹ (*Outils de Traitement de l'Information Multimodale*). Ce projet avait pour ambition de développer des outils permettant de rendre disponibles et exploitables des corpus de parole multimodaux. Le projet OTIM était fortement interdisciplinaire : il rassemblait des spécialistes de la description et du traitement de chacune des modalités impliquées dans l'interaction multimodale (la phonétique, la phonologie, la syntaxe, la sémantique, le discours ou la pragmatique et les gestes). Ce projet avait un double objectif linguistique et technologique.

L'idée générale de ce projet, à laquelle j'adhère pleinement, propose que l'analyse linguistique doit tenir compte des différentes modalités d'expression de l'information et passe par la description de données qui illustrent les interactions particulières entre les différents domaines et modalités (interactions parole-gestes-syntaxe, interactions sémantique-pragmatique-prosodie, etc., figure 2.a). De telles descriptions ne peuvent être menées qu'à l'aide de corpus annotés, dans lesquels toutes ces informations sont décrites.

¹ "Outils de Traitement de l'Information Multimodale" (OTIM), ANR 2008-2011, Responsable scientifique: Philippe Blache, LPL. (<http://aune.lpl-aix.fr/~otim/>)

Cette annotation multi-niveaux est fondamentale pour le niveau segmental. La structure prosodique est souvent mise en relation avec des unités linguistiques de haut niveau, comme la syntaxe, la sémantique ou le discours. Ces unités couvrent un domaine temporel large correspondant bien à l'empan des structures prosodiques. En revanche, l'empan temporel correspondant aux segments phonétiques est très restreint. L'étude des variations affectant ces segments est ainsi très souvent cantonnée au cadre syllabique ou lexical. Nous espérons ainsi pouvoir mettre en relation certains phénomènes de variation segmentale avec des structures linguistiques de haut niveau.

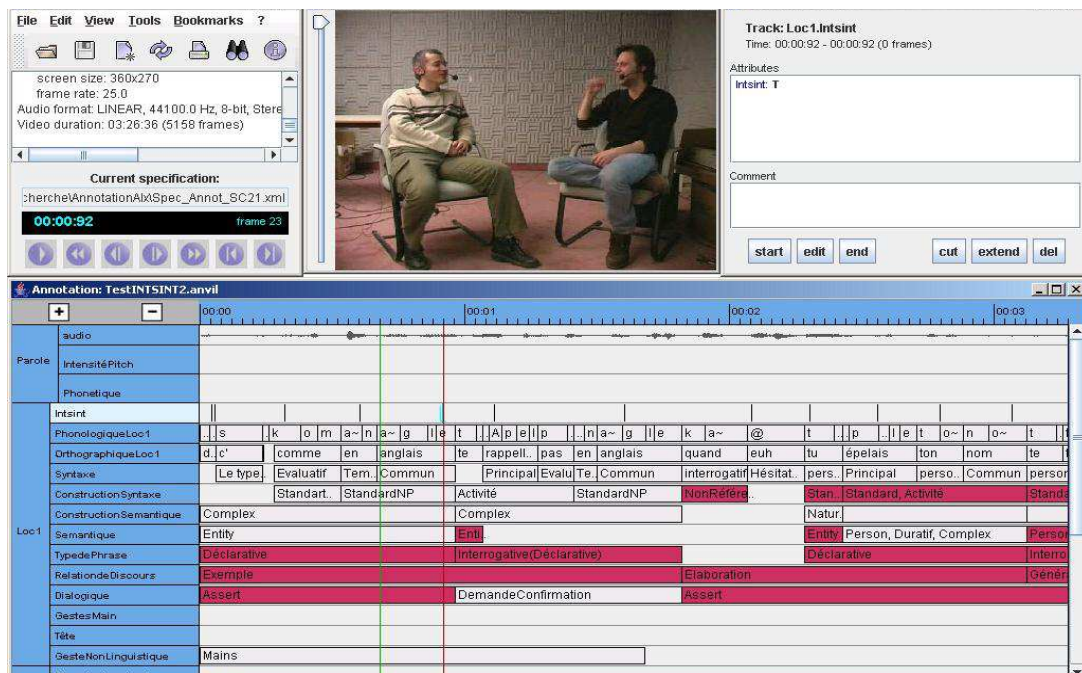


Figure 2.a: Exemple d'annotation multi-niveaux sous ANVIL: annotation gestuelle et représentation des données (aimablement transmise par Roxane Bertrand)

Le développement des outils s'est formé autour du corpus CID (*Corpus of Interactional Data*), élaboré au Laboratoire Parole et Langage. Ce corpus est un outil multimodal contenant 8h de conversation en binômes (16 locuteurs, dont 6 hommes et 10 femmes). De nombreux chercheurs du LPL ont été mobilisés autour de l'annotation mais aussi de l'analyse de ce corpus. Il présente en effet la particularité d'une parole interactive familière permettant d'observer des phénomènes linguistiques spécifiques et, en ce qui concerne la parole, de nombreux phénomènes de **réduction**. Ainsi, l'exploitation des annotations et outils développés dans le cadre du d'OTIM constitueront la base de mes travaux sur les phénomènes de réduction chez le locuteur sain.

Contribution: Annotation syllabique

Bigi, B., Meunier, C., Nesterenko, I., Bertrand, R. (2010) "Annotation automatique en syllabes d'un dialogue oral spontané", Actes des Journées d'Etude sur la Parole, Mons (Belgique), mai 2010, 181-184.

Bigi, B., Meunier, C., Nesterenko, I., Bertrand, R. (2010) "Syllable Boundaries Automatic Detection in Spontaneous Speech", Proceedings of Language Resource and Evaluation Conference, Malte, mai 2010, 3285-3292.

Dans le cadre d'OTIM, Brigitte Bigi a développé un outil d'annotation automatique de la parole composé de plusieurs modules¹ (Bigi, 2012). J'ai ainsi pu participer au développement d'un module d'annotation syllabique du corpus CID. Mon rôle a consisté en l'élaboration de critères permettant de déterminer les frontières syllabiques. Cette annotation est basée sur la chaîne de phonèmes issue de la phonétisation du corpus sans tenir compte des frontières lexicales. Le système développé consiste à définir un ensemble de règles de segmentation entre phonèmes, répartis en six classes:

- V-Voyelles: i e ε a α ɔ o u y ø œ ə ĩ ã õ õ
- G-Glides: j ɥ w
- L-Liquides: l ʀ
- O-Occlusives: p t k b d g
- F-Fricatives: f s ʃ v z ʒ
- N-Nasales: m n ŋ ɲ

Deux principes généraux permettent de cadrer les règles de syllabation: 1/ une syllabe contient une seule voyelle; 2/ une pause est une frontière de syllabe (les pauses silencieuses supérieures à 200 ms ont été annotées automatiquement). L'annotation syllabique constitue donc en une recherche de frontières de syllabes entre deux voyelles.

Nous avons organisé les règles en deux types : des **règles générales**, applicables à tout type de situation, et des **règles exceptions** qui peuvent être modulées selon le type de corpus (dialogue spontané ou non, etc.) (table 2.a).

Règles générales				☐	Règles exception				☐
Séquence	Règle	Exemple	☐	Séquence	Règle	Exemple	☐		
1	VV	V.V	poète	po.ɛt	1	VXGV	V.XGV	spécial	spe.sjal
2	VXV	V.XV	limité	li.mi.te	2	VFLV	V.FLV	à vrai dire	a.vrɛ.dir
3	VXXV	VX.XV	comme ça	kom.sa	3	VOLV	V.OLV	il trouve	i.truv
4	VXXXV	VX.XXV	avec moi	a.vek.mwa	4	VFLGV	V.FLGV	effroyable	e.frwa.jabl
5	VXXXXV	VX.XXXV	il se promène	il.spro.mɛn	5	VOLGV	V.OLGV	à trois	a.trwa
6	VXXXXXV	VXX.XXXV	alors je crois	a.lorz.krwa	6	VOLOV	VOL.OV	capable de	ka.pabl.d

Table 2.a: Règles générales (à gauche) et règles d'exception (à droite) pour l'annotation en syllabes des suites de phonèmes.

¹ <http://www.lpl-aix.fr/~bigi/sppas/>

La première règle générale applique le principe 1 selon lequel il n’y a qu’une voyelle par syllabe. La deuxième règle reflète la tendance universelle à privilégier les syllabes ouvertes. Les règles générales 4, 5 et 6 satisfont la règle du *Maximum Onset Principle* (Selkirk, 1981) pour laquelle, dans un groupe de consonnes intervocaliques, le maximum de consonnes doit être attribué à l’onset de la seconde syllabe plutôt qu’à la coda de la première. La troisième règle doit être considérée en fonction des “règles exceptions” 1, 2 et 3, dans lesquelles les clusters intervocaliques sont constitués de deux consonnes. Ces règles montrent que pour un cluster de deux consonnes le “Maximum Onset Principle” doit être appliqué prudemment et en fonction du principe de sonorité, c’est-à-dire de la nature des consonnes du cluster. Ainsi, la règle générale ne s’applique que lorsque le principe de sonorité est violé au sein du cluster. Ces règles permettent d’obtenir une annotation syllabique de plusieurs niveaux alignée sur le signal de parole (figure 2.b). Quelques analyses effectuées sur le corpus du CID montrent que la très grande majorité des découpages syllabiques concernent les règles 2 et 3.

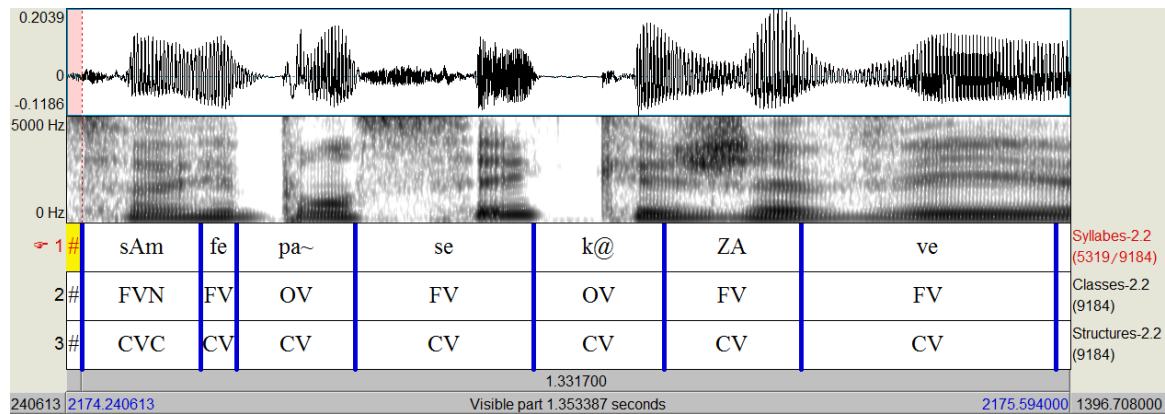


Figure 2.b: Exemple de syllabation automatique sur le corpus du CID (locuteur AB) obtenues avec SPASS. Séquence: *Ca me fait penser que j'avais euh produit* /samfɛpɑ̃sekɑ̃zɑvɛə/. Les frontières sont issues de l'alignement automatique. La Tier 1 correspond aux syllabes phonétiques, la Tier 2 aux syllabes en macro-classes et la Tier 3 à la structure syllabique (Consonne/Voyelle).

L'intérêt de ce niveau d'annotation sur un corpus de parole continue non contrôlée est double. Dans un premier temps, il permet d'obtenir des statistiques sur la structure et la fréquence des syllabes du français en parole non contrôlée après la réorganisation syllabique (re-syllabation) qu'entraîne la parole continue. Ensuite, disposer de cette annotation permet d'envisager les analyses segmentales (structures acoustiques des phonèmes) et suprasegmentales (organisation rythmique et accentuelle du discours) en fonction du découpage syllabique. Par exemple, cette annotation permettra de mettre en perspective le rôle et la hiérarchie des frontières lexicales et/ou syllabiques dans la coarticulation.

Contribution: **Annotation et segmentation phonétique**

Meunier, C., Nguyen, N. (2013) "Traitement et analyse du signal de parole", Chapitre 2 in Méthodes et outils pour l'analyse phonétique des grands corpus oraux, NGUYEN Noël, ADDA-DECKER Martine, Traité IC2, série Cognition et traitement de l'information, Hermès, pages 85-119. (ISBN : 9782746245303).

Si une seule publication est reliée à cette partie, elle n'en reste pas moins un axe majeur de mon parcours en phonétique. D'une certaine façon, peu de choses sont à dire sur l'annotation phonétique mais tous mes travaux partent de là. Ce travail de balisage récurrent du signal de parole est une constante depuis mon doctorat. Cette expertise est probablement là où je situerais mes compétences les plus affirmées. De cette expérience se dégage une sorte de devise: "voir le signal de parole pour le comprendre". Evidemment, l'annotation phonétique a connu, depuis une vingtaine d'années un changement fondamental et décisif. L'utilisation de grands corpus de parole ainsi que les outils développés pour leur exploitation ont rendu quasiment obsolète l'expertise manuelle en ce qui concerne les langues déjà bien documentées. Il est probable que le travail initial de Segmentation Manuelle évolue désormais vers un travail de Correction Manuelle de l'Alignement (voir ci-après la mise en perspective des annotations automatiques et manuelles). Dans un avenir proche, il est donc probable que très peu (ou pas) de corpus seront encore segmentés manuellement dans leur ensemble. Toutefois, le risque est que l'utilisation exclusive des résultats de l'alignement automatique entraîne une certaine *cécité* vis-à-vis du signal de parole et rende difficile l'identification de phénomènes liés au style et à la nature de la parole observée (réduction, fusion, métoplasmes, etc.). Pour cette raison, il nous semble indispensable de *regarder* le signal de parole même si les outils actuels pourraient nous en dispenser.

Les étapes actuelles permettant l'annotation phonétique des corpus sont les suivantes:

1. transcription du signal de parole
2. conversion graphème-phonème de la transcription
3. alignement temporel des phonèmes sur le signal de parole
4. Correction éventuelle de l'alignement.

Seule l'étape 4 concerne totalement le phonéticien, même si les conventions de transcription (étape 1) peuvent nécessiter ses compétences. Nous voyons ainsi que l'expertise du phonéticien peut devenir optionnelle. Toutefois, nous verrons ici et dans la partie suivante pourquoi les incursions dans le signal sont toujours indispensables.

La transcription

Si la transcription d'une parole contrôlée (discours, journal télévisé, etc.) pose rarement problème, ça n'est pas le cas pour un style de parole relâché ou déviant. En effet, il arrive que

certaines parties du signal ne soient pas intelligibles (c'est relativement rare en parole spontanée, mais assez fréquent en parole pathologique). Par ailleurs, transcrire une parole atypique, quelle soit normale ou pathologique, pose le problème de la correspondance entre le codage choisi et la production des locuteurs. En effet, le code orthographique n'est pas adapté aux phénomènes de variations, de réductions, d'hypoarticulation, etc. D'une certaine façon, une transcription totalement phonétique serait la plus adaptée à ce qui est effectivement produit. Or, à ma connaissance, elle n'est pas utilisée pour les grands corpus pour plusieurs raisons: 1/ elle n'est utilisable que par des experts phonéticiens; 2/ elle est chronophage et source d'erreurs (une connaissance parfaite de l'alphabet phonétique est indispensable); 3/ elle rend impossible l'accès à d'autres niveaux annotation, morpho-syntaxique, notamment.

Ainsi, le plus souvent, une transcription orthographique est utilisée. Elle peut ainsi être strictement orthographique (TO, Blanche-Benveniste, 1999 & Adda-Decker et al., 2008) ou enrichie (TOE, Bertrand et al., 2008) Les deux types de transcription présentent divers avantages et inconvénients:

- Transcription Orthographique
 - Avantages: rapidité, simplicité, pas d'experts requis
 - Inconvénients: plus d'erreurs d'alignement, pas de codage des phénomènes de variation ni des phénomènes spécifiques à la parole spontanée (élisions, disfluences, etc.)
- Transcription Orthographique Enrichie
 - Avantages: moins d'erreurs d'alignement, codage et information sur les variations perçues
 - Inconvénients: chronophage, complexité d'utilisation (source d'erreurs et d'hétérogénéité dans les transcriptions), hétérogénéité des transcriptions (plus le codage est complexe, plus les transcripateurs peuvent s'en éloigner ou faire des erreurs)

Evidemment la TOE peut être plus ou moins complexe selon le degré de précision de l'enrichissement à donner et selon la difficulté d'utilisation du codage. Pour le projet TYPALOC, nous avons choisi d'utiliser une version simplifiée du codage utilisé pour le projet OTIM (transcription du corpus CID). Ainsi, seuls les phénomènes pertinents pour ce projet étaient conservés (voir *Conventions de Transcription TYPALOC* en Annexe).

En général, les phénomènes connus de variations de la parole sont intégrés dans les aligneurs (Adda-Decker et al., 2013). Ainsi, les élisions du schwa, les variations dialectales possibles, les métaplasmes les plus connus (*je sais* produit /ʃɛ/) sont pris en compte dans la procédure automatique ce qui réduit considérablement les erreurs d'alignement et fournit ainsi une annotation phonétique tout à fait acceptable.

Toutefois, pour le phonéticien, la TOE présente des informations pertinentes. Je veux parler du seuil de perception des variations produites. En effet, les formes de variations présentes dans le signal de parole sont plus ou moins saillantes pour l'auditeur: l'élision du schwa (figure 2.c), les métaplasmes stéréotypés (*je suis produit /ʃɥi/*) sont en général identifiés par les transcrip-teurs utilisant la TOE. En revanche, certaines omissions ou phénomènes de coalescence sont opaques et imperceptibles pour les transcrip-teurs. Ainsi, de façon indirecte, la TOE fournit un inventaire des phénomènes de variation perçus. Une analyse des corpus peut ensuite nous fournir les variations non perçues. Cette distinction est importante car nous faisons l'hypothèse que ces deux types de phénomènes (saillants ou opaques) relèvent de processus et de dépendances linguistiques différents.

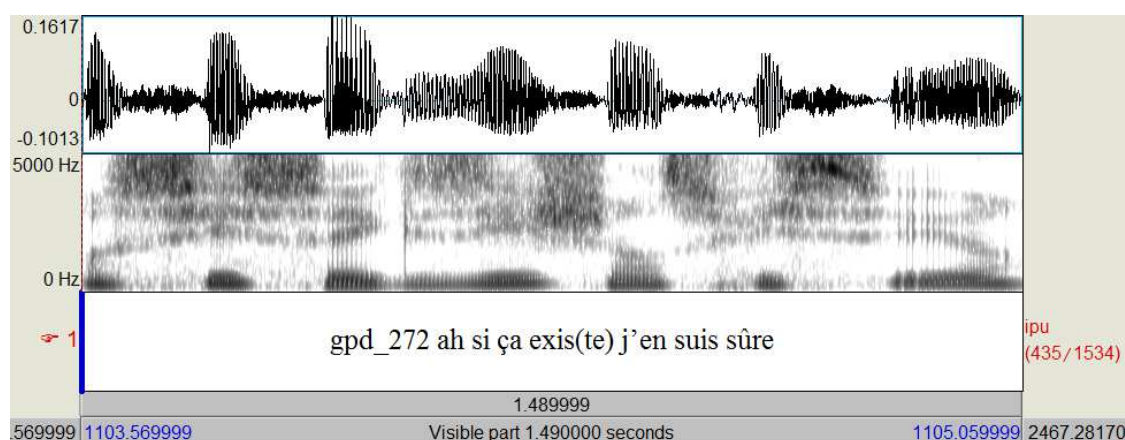


Figure 2.c: exemple simple de TOE dans laquelle l'omission est perçue et notée entre parenthèses. Corpus CID (locuteur AB), conventions OTIM. Séquence produite: "ah si ça existe, j'en suis sûre" /asisaegziszãstqisyɾ/.

Correction de l'alignement ou segmentation

Mon expérience de la segmentation du signal de parole m'a permis d'établir des critères ainsi qu'un cadre permettant de réaliser au mieux soit une segmentation du signal de parole, soit une correction de l'alignement (Meunier and Nguyen, 2013). Il s'agit d'un outil permettant de baliser le signal de parole pour toute personne (chercheurs, étudiants, etc.) qui en aurait l'utilité. Il s'agit également d'une réflexion sur la nature du signal acoustique et la relation entre représentations phonétiques abstraites et discontinuités du signal de parole.

Notons que *segmentation*, *alignement* et *annotation* sont souvent employés pour désigner un même résultat : le marquage temporel des segments phonétiques positionnés sur le signal de parole. Ces termes relèvent toutefois de processus différents pour des raisons historiques et disciplinaires. La *Segmentation Manuelle* (SM) désigne le travail d'expertise humaine consistant à marquer les frontières de segments phonétiques par des étiquettes de début et de fin. Le terme

Alignement Automatique (AA) est issu du traitement automatique et consiste, par le biais de modèles acoustiques, à faire correspondre une suite de symboles phonétiques (issue d'une transcription) avec le signal de parole. Enfin, l'*Annotation Phonétique* (AP) fait référence aux résultats soit de la segmentation, soit de l'alignement. Il est utilisé en référence à d'autres niveaux d'annotation comme les niveaux morpho-syntaxique, prosodique, discursif, voire même gestuel pour le signal vidéo.

Segmenter le signal revient à apposer sur le signal de parole des étiquettes temporelles représentant approximativement le début ou la fin d'une unité linguistique. C'est évidemment dans le terme 'approximativement' que se trouve tout le débat sur la segmentation de la parole, tout particulièrement en ce qui concerne les unités phonétiques. L'ensemble de nos connaissances sur la coarticulation (Farnetani, 1997) et sur la propagation des traits phonétiques sur de longues séquences (Nguyen et al., 2004) a montré qu'il est illusoire de vouloir représenter les segments phonétiques comme une suite d'éléments contenus entre deux frontières. S'il est indéniable que la production des phonèmes est ordonnée dans le temps, les multiples articulateurs en jeu et leurs mécanismes complexes de synchronisation et d'anticipation suggèrent que la sortie acoustique observable comporte de multiples indices répartis au-delà de la manifestation saillante du phonème sur le signal de parole. Cette relation complexe entre niveaux linguistique, articuloire et acoustique nous rappelle que les frontières de segments ne doivent pas être confondues avec des frontières de phonèmes (Fant, 1973). En effet, la notion abstraite de *phonème* ne peut être mise en relation directe avec des discontinuités du signal acoustique (Rossi, 1990). On admet donc que les étiquettes de début et de fin des unités phonétiques, qu'elles soient apposées automatiquement ou manuellement, ne représentent pas des frontières réelles de phonèmes. Il s'agit juste de repères permettant de procéder à des analyses sur le signal de parole. Car la segmentation n'est pas un but en soi, mais plutôt un outil qui doit donc répondre à des questions préalables. La plupart du temps la segmentation sert à délimiter des segments au travers desquels des analyses acoustiques seront menées. Les étiquettes de début et de fin vont alors permettre d'extraire des durées, des valeurs formantiques, la fréquence fondamentale, l'énergie, etc., ces mesures étant prises sur la totalité ou sur un point précis du segment.

Pourquoi segmente-t-on la parole? En premier lieu pour maîtriser le flux continu de la parole, difficilement appréhendable tel quel. Ensuite pour extraire des données acoustiques (durées, spectre, F0, etc.). Sans un balisage, quel qu'il soit, il n'est pas possible d'analyser la parole pour en faire un produit interprétable. La segmentation de la parole est également nécessaire pour de nombreuses expériences de perception dans lesquelles une partie du signal de parole doit être

extraite. Enfin, ce balisage est également indispensable pour l'évaluation des outils d'alignement automatique.

Nous avons ainsi élaboré des **critères de segmentations** entre macro-classes phonétiques de façon à rendre accessible cette tâche parfois ardue pour les non spécialistes. Il s'agit de fournir des indices utilisés très couramment pour caractériser des frontières de segments que chacun pourra utiliser selon leur pertinence avec les objectifs posés. La segmentation se fait en deux temps: en premier un repérage global des macro-classes (forme globale des segments), ensuite un repérage approximatif des zones frontières (zone de transition entre deux segments). Le choix d'une position frontière est ensuite dépendant du type de transition rencontrée:

- Présence d'une discontinuité sur la zone frontière : l'étiquette est posée sur cette discontinuité
- Chevauchement des indices : a/ primauté d'un indice ou b/ étiquetage au centre du chevauchement
- Transition graduelle : étiquetage au centre de la transition

Les critères de segmentations sont ensuite déclinés selon le type de macro-classes en contact.

Quatre macro-classes sont distinguées:

- Occlusives /p t k b d g/
- Fricatives /f s ʃ v z ʒ ʁ/
- Consonnes vocaliques /l ʀ m n j ŋ w/
- Voyelles /i y u e ø o ε œ ɔ a α ã õ œ ã/

Ces regroupements sont justifiés par le type de zone transitoire caractéristique de chaque segment phonétique. Ainsi, la classe des consonnes vocaliques comporte des consonnes très variées d'un point de vue articulatoire mais qui présentent des transitions semblables dans la mesure où toutes ces consonnes sont caractérisées par une structure de formant. Ainsi pour chaque type de combinaison des indices acoustiques pertinents permettent d'identifier une zone frontière.

Contribution: **L'annotation phonétique: automatique ou manuelle?**

Bertrand, R., Blache, P., Espesser, R., Ferré, G., **Meunier**, C., Priego-Valverde, B., Rauzy, S. (2008) "Le Cid - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle", *Traitement Automatique des Langues*, 49, 105-134.

Meunier, C., Nguyen, N. (2013) "Traitement et analyse du signal de parole", Chapitre 2 in *Méthodes et outils pour l'analyse phonétique des grands corpus oraux*, NGUYEN Noël, ADDA-DECKER Martine, *Traité IC2, série Cognition et traitement de l'information*, Hermès, pages 85-119. (ISBN : 9782746245303).

Fougeron, C., Audibert, N., Fredouille, C. **Meunier**, C. Gendrot, C., Panseri, O. (2010) "Comparaison d'analyses phonétiques de parole dysarthrique basées sur un alignement manuel et un alignement automatique", *Actes des Journées d'Etude sur la Parole, Mons (Belgique)*, mai 2010, 365-368.

Audibert, N., Fougeron, C. Fredouille, C. **Meunier**, C. (2010) "Evaluation d'un alignement automatique sur la parole dysarthrique", *Actes des Journées d'Etude sur la Parole, Mons (Belgique)*, mai 2010, 353-356.

Avant l'arrivée des systèmes d'alignement automatique, les corpus de parole était segmentés à la main, par des experts. Ce travail, coûteux, était toutefois possible dans la mesure où les corpus enregistrés étaient de taille restreinte. Cette expertise ne nécessitait pas de transcription préalable, dans la mesure où la plupart des corpus correspondaient à de la lecture de pseudo-mots, de mots, de phrases ou de textes. Les analyses phonétiques, s'intéressent désormais à des types de parole très variés et moins prototypiques. La forte variabilité présente dans ces productions non contrôlées, et souvent contextualisées, nécessite un très grand nombre de données de façon à pouvoir extraire des analyses pertinentes d'un point de vue statistique. Cette masse de données rend donc impossible une annotation simplement manuelle du signal de parole.

Ainsi, a-t-on encore besoin d'experts humains pour annoter le signal de parole? La question est évidemment pertinente et d'actualité au moment où l'analyse des grands corpus de parole est en plein essor et qu'il est difficilement envisageable de segmenter « à la main » ce type de corpus, essentiellement en raison de l'aspect chronophage¹ de telles procédures. Par ailleurs, la segmentation manuelle donne-t-elle de meilleurs résultats que l'alignement automatique ? Certaines études (Cosi et al., 1991; Pitt et al., 2005; Wesenick and Kipp, 1996) ont montré de nombreuses irrégularités et différences dans les annotations produites par plusieurs experts humains, suggérant ainsi qu'un outil automatique serait plus fiable ou, tout au moins, plus régulier que l'expertise humaine. Il est vrai que l'utilisation et l'amélioration des outils d'alignement automatique (AA) ont considérablement changé le travail de segmentation manuelle et les deux approches, manuelle et automatique, doivent désormais s'inscrire dans une dynamique de complémentarité plutôt que d'opposition.

¹ On estime qu'il faut 1h pour annoter phonétiquement 1min de signal de parole. Cette estimation est toutefois très approximative et sous-estimée selon la précision du travail effectué.

Mais la question de l'Annotation automatique ou manuelle ne se pose pas réellement en termes de choix. Les analyses phonétiques portant sur de très grands corpus se basent désormais le plus souvent sur les résultats de l'alignement automatique, dans la mesure où il n'est pas envisageable d'annoter manuellement plusieurs centaines de milliers de phonèmes. De ce fait, l'expertise manuelle se raréfie sur ce type de corpus et la question est désormais de savoir si l'Alignement Automatique nous fournit un balisage du signal de parole suffisamment fiable pour y effectuer des analyses phonétiques de qualité (Machac and Skarnitzl, 2009). Une question subsidiaire serait ensuite de savoir quels types d'analyse phonétique sont envisageables avec l'AA et lesquels ne le sont éventuellement pas. Pour répondre à ces questions, les résultats de l'Alignement Automatique ont été corrigés par des experts humains (Correction Manuelle de l'Alignement, CMA) de façon à faire apparaître les avantages et inconvénients respectifs des deux approches.

L'avantage en termes de coût et de rapidité revient évidemment à l'AA qui permet l'annotation phonétique de très grand corpus oraux, ce qui est quasiment impossible à envisager pour la segmentation manuelle. Concernant la qualité et la régularité de l'annotation produites, les avis sont partagés. On considère souvent que l'expertise manuelle est plus précise et qu'elle permet une certaine adaptabilité en fonction des productions réelles des locuteurs. Toutefois, certains travaux (Pitt et al., 2005; Wesenick and Kipp, 1996) ont pu mettre en évidence une forte variabilité inter experts. Ces résultats ne sont pas surprenants lorsque les experts sont libres d'utiliser les critères de leur choix. En revanche, lorsque les experts s'accordent des critères de segmentation communs, les écarts sont assez faibles (entre 1 et 3 ms en moyenne; Volin et al. 2008 cité par Machac and Skarnitzl, 2009). Ainsi, l'écart inter-expert ne serait pas forcément dû à une localisation approximative des frontières de phonèmes (même si cela reste possible), mais plutôt à des choix différents concernant l'affectation des indices acoustiques pertinents pour l'identification des zones frontières.

Nous avons pu effectuer une comparaison entre AA et CMA en français dans le cadre de plusieurs projets portant sur de large corpus de parole peu prototypique ou altérée (parole conversationnelle¹, d'une part, et parole dysarthrique², d'autre part). Notons toutefois, que la segmentation manuelle, dans ce cas, correspond à une correction de l'alignement (déplacement, insertion ou retrait des étiquettes de l'AA) et non à un processus de segmentation, ce qui est

¹ Dans le cadre du projet ANR OTIM (P. Blache).

² Dans le cadre du projet ANR DesPhoAPaDy (C. Fougeron).

assez différent. Dans les deux cas, des décalages ont pu être observés entre les étiquettes de l'AA et celles issues de la correction de l'alignement. Notamment, les travaux sur la parole conversationnelle (Bertrand et al., 2008) font apparaître une sous-estimation systématique des segments vocaliques pour l'AA se traduisant par un marquage plus tardif du début des voyelles ainsi qu'un marquage plus précoce de la fin (figure 2.d).

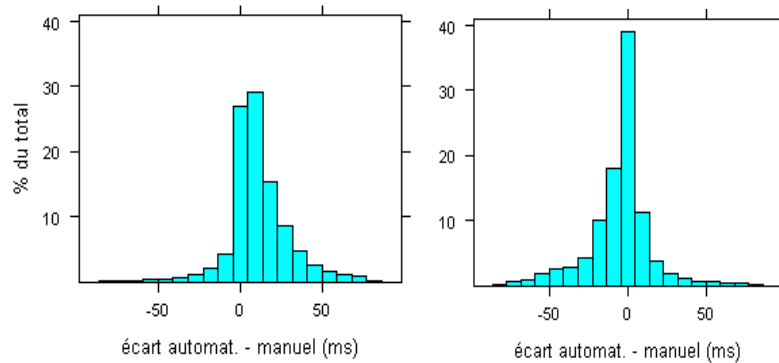


Figure 2.d: Histogramme des écarts en début de voyelle (à gauche) et en fin de voyelle (à droite). L'écart médian automatique/manuel est de 9ms au début de la voyelle, de 0ms à la fin et de 3ms au milieu de la voyelle.

Concernant la parole dysarthrique (Audibert et al., 2010), cette évaluation s'est faite en comparant des outils d'alignement automatique et des alignements effectués par des experts humains. L'analyse menée met en évidence une variabilité des performances du système d'alignement en fonction de la sévérité de la dysarthrie, mais également une différence entre experts humains. Nous avons alors cherché à savoir si cette variabilité avait des conséquences sur les résultats des analyses acoustiques effectuées sur les patients. Ce ne fut pas le cas pour l'ensemble des analyses qui prennent en compte la globalité d'un segment (analyse des voyelles, centre de gravité des fricatives, Fougeron et al., 2010). Il ressort de ces résultats que l'application d'une procédure d'alignement automatique pour l'analyse de critères phonéto-acoustiques sur la parole dysarthrique est tout à fait envisageable, même si, en fonction des critères étudiés, elle nécessitera une vérification manuelle.

A travers ces résultats, il est possible d'entrevoir la complémentarité entre AA et CMA en vue d'une exploitation pour des analyses phonétiques. Lorsque les analyses phonétiques portent sur l'ensemble des segments et notamment sur leur durée globale ou sur leur tenue (extraction de formants au centre d'une voyelle ou *Centre de Gravité* du bruit d'une fricative, par exemple), l'annotation fournie par l'AA semble satisfaisante et ne nécessite pas de correction manuelle. En revanche, lorsque les analyses portent sur des zones transitoires dont les localisations doivent être précises (événements acoustiques tels que l'explosion d'une plosive, des zones d'assimilation de voisement ou des phénomènes de coarticulation), l'annotation fournie par l'AA

apparaît trop approximative. Ainsi, en fonction des besoins des analyses, l'AA sera utilisable tel quel ou devra être corrigé par des experts. L'AA rend possible de larges études quantitatives portant sur de très grands corpus de parole dès lors que les analyses portent sur des segments dont les zones frontières peuvent-être approximatives. En revanche, des études portant sur des segments très courts et précis sont pour l'instant limités à une faible quantité de données dès lors qu'elles nécessitent une correction manuelle.

Enfin, la correction manuelle semble indispensable dans les cas d'erreur importante, c'est-à-dire, lorsque l'aligneur propose une annotation déconnectée du signal de parole (figure 2.e). Cela peut être le cas lorsque la transcription est trop éloignée de la réalisation réelle du locuteur, ou encore lorsque la production du locuteur est très 'altérée' (ce qui est le cas chez des patients dysarthriques, mais aussi en parole naturelle lors de production très fortement hypo-articulées).

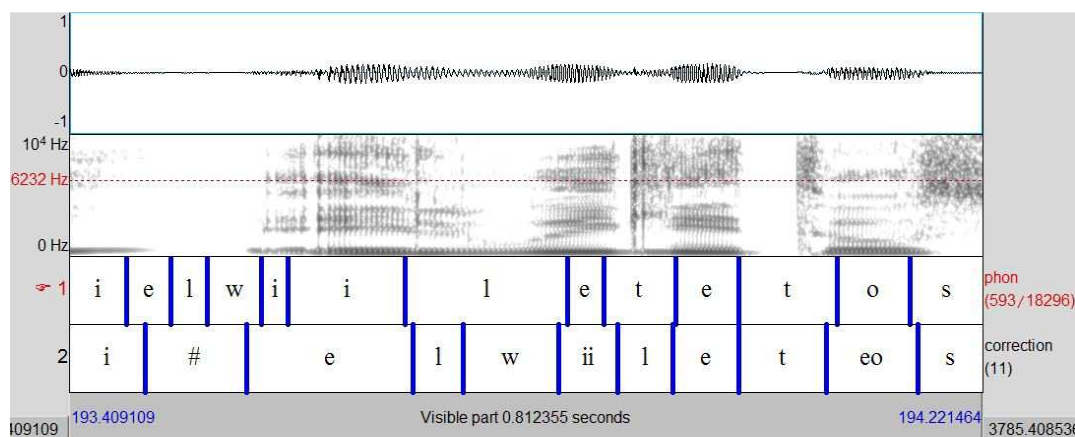


Figure 2.e: Séquence "[...] et Louis il était au [...]" (extrait du CID). Exemple d'erreurs massives de l'aligneur ("phon") comparées avec la correction d'un expert ("correction"). Les étiquettes de phonèmes ne sont plus en phase avec la réalisation effective.

Il apparaît ainsi que l'annotation phonétique manuelle du signal de parole, en tant que tâche isolée, disparaîtra probablement en raison de la fiabilité des données fournies par les aligneurs et du gain de temps. Toutefois cette évolution n'est pas anodine dans la mesure où la correction manuelle de l'alignement se fait toujours en référence à l'annotation de l'aligneur et, même si l'expert utilise ses propres critères, ses choix sont influencés par la localisation des étiquettes déjà présentes sur le signal. Notamment, cette approche favorise une représentation en segments discrets (phonèmes balisés par une étiquette de début et de fin) qui peut être éloignée ou, du moins, peu informative par rapport aux réalisations effectives des locuteurs. De plus, dans la pratique de la correction de l'alignement, l'expert va éviter de déplacer une étiquette lorsqu'il doute (frontière ambiguë ou chevauchement) ou lorsque la correction ne porte que sur quelques millisecondes. Cette situation tend à privilégier l'alignement automatique au détriment des choix de l'expert. Mais l'influence de l'alignement sur

l'annotation des experts humains peut entraîner une régularité dans les pratiques qui finalement peut être souhaitable. En retour, le regard des experts humains sur l'alignement automatique et ses imperfections peut permettre une amélioration de ces outils au travers d'une meilleure définition des modèles acoustiques utilisés. Notamment, l'identification de phénomènes spécifiques à certains types de parole doit être exploitée dans le développement des aligneurs. On peut ainsi envisager une complémentarité entre les approches manuelles et automatiques à la fois dans leur exploitation actuelle, mais aussi en vue d'une amélioration des techniques et des connaissances sur la langue parlée.

Contribution	Annotation de la parole atypique
	<p>Meunier, C., Nguyen, N. (2013) "Traitement et analyse du signal de parole", Chapitre 2 in Méthodes et outils pour l'analyse phonétique des grands corpus oraux, NGUYEN Noël, ADDA-DECKER Martine, Traité IC2, série Cognition et traitement de l'information, Hermès, pages 85-119. (ISBN : 9782746245303).</p>
	<p>Meunier, C. (2013) "Phoneme deletion and fusion in conversational speech", ExAPP 2013 Experimental Approaches to Perception and Production of Language Variation, Copenhagen (Denmark), mars 2013</p>
	<p>Laaridh I. & Fredouille C. & Meunier C. (soumis) "Automatic detection of phoneme-based anomalies in dysarthric speech», Special Issue of ACM Transactions on Accessible Computing (TACCESS) on Speech and Language Interaction for Daily Assistive Technology (SLPAT), (article soumis en avril 2014).</p>

Pourquoi parle-t-on de parole atypique? Y'a-t-il une parole "typique"? C'est dans le cadre de ces questions que le projet ANR TYPALOC a été élaboré. Ce terme peut, en premier lieu, concerner la parole spontanée. Cela peut paraître quelque peu absurde d'utiliser cet adjectif pour un type de parole tout à fait courant. En fait, ce qui est considéré comme atypique ici, ce n'est pas le style de parole lui-même, mais plutôt les réalisations phonétiques très éloignées des descriptions standard (généralement établies à partir de parole lue, préparée).

Le terme atypique concerne également la parole produite par des locuteurs atteints de pathologie de la parole. Dans ce cas, le type de parole, comme les réalisations sont considérées comme atypiques.

Dans le cas de la dysarthrie, les phénomènes atypiques observés sont très similaires à ceux présents en parole spontanée (réductions, fusions, approximations, omissions, allongements, etc.). Il devient alors délicat de distinguer ce qui est caractéristique du style de parole de ce qui est déviant ou pathologique (figure 2.f).

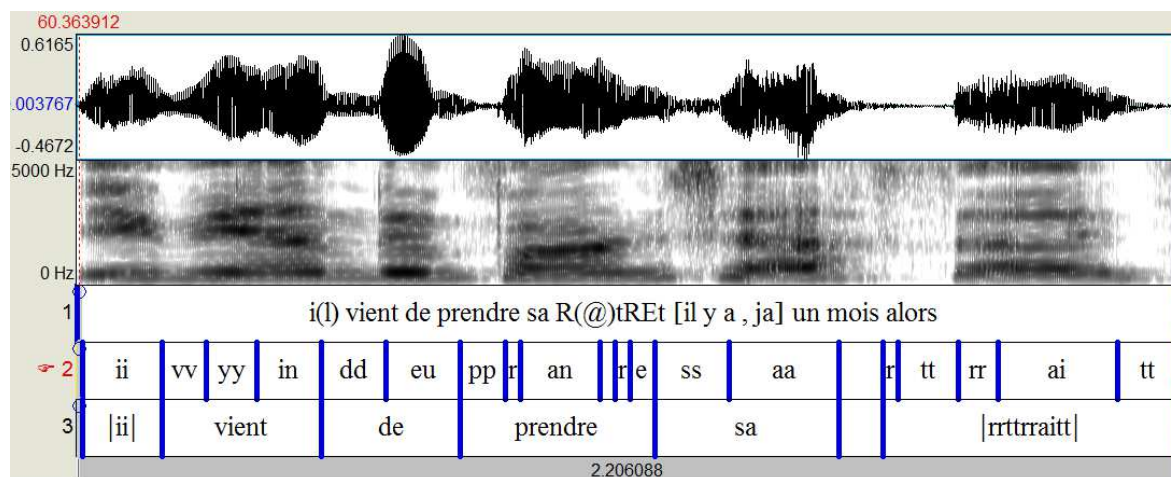


Figure 2.f: production spontanée d'une locutrice atteinte de Sclérose Latérale Amyotrophique (PHO-000814-01_P01): séquence "Il vient de prendre sa retraite..." produit /ivjɛdpãdsagʁɛt/. Toutes ces réductions sont possibles en parole spontanée. Zones de réduction identifiées: il produit /i/, élision du /ə/. En revanche, l'élision du /R/ dans *prendre* est possible en spontané mais non attendu dans cette séquence; Par ailleurs le mot retraite /ʁtʁɛt/ est extrêmement réduit et cela est perceptible.

L'annotation des phénomènes atypiques est indispensable pour le travail des phonéticiens. Je ne parle pas là spécifiquement de voyelles centralisées ou de consonnes dévoisées ou réalisées approximantes. Tous ces cas sont analysables tels quels, automatiquement, du moment que les phonèmes sont alignés. Je veux plutôt parler de réalisations atypiques non identifiés par les transpositeurs et, de fait, alignés de façon inappropriée (2.g).

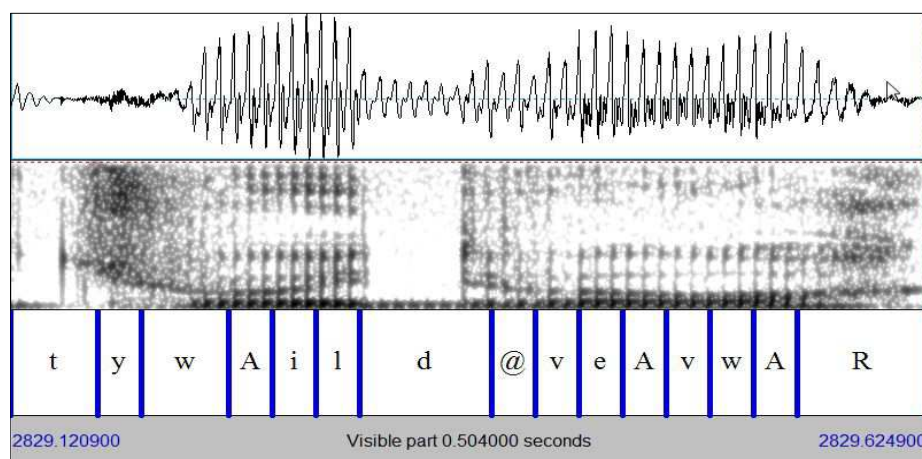


Figure 2.g: Exemple de fusion *tu vois il devait avoir*, transcrit /tywAild@veAvwaR/. La réalisation effective est bien en deçà de la transcription. Notamment la séquence /@veAvwa/ est insegmentable et ne peut pas être représentée sous la forme d'une suite d'unités discrètes dans laquelle certaines unités seraient réalisées et d'autres non.

Que ce soit en parole spontanée ou en parole pathologique, ces phénomènes posent des problèmes certains concernant leur annotation. Ils ne cadrent pas avec notre représentation paradigmatique ou syntagmatique de l'annotation phonétique. Il s'agit de phénomènes de coalescence pour lesquels il est difficile de distinguer ce qui est produit de ce qui ne l'est pas.

Une solution, pour l'instant, consiste à les identifier et les baliser en spécifiant leur caractéristique de fusion. Toutefois, ce processus n'est pas simple car il implique que l'ensemble du signal de parole soit inspecté de façon à répertorier tous ces phénomènes.

C'est dans ce cadre qu'Imed Laaridh a commencé son travail de thèse en 2013 sous la codirection de Corinne Fredouille et de moi-même. Dans la poursuite des travaux menés par Corinne Fredouille (Fredouille and Pouchoulin, 2011) sur la parole dysarthrique, I. Laaridh développe un outil de détection automatique des *déviances*. L'objectif de ce travail est d'étudier comment les outils du traitement automatique de la parole pourraient permettre de traiter de larges populations de patients dysarthriques et de focaliser l'attention des experts humains sur des zones de déviance bien identifiées du signal en vue d'analyses plus fines. La collaboration entre chercheurs en traitement automatique et phonéticiens pourrait apporter ici de nombreuses avancées. Par ailleurs, dans la mesure où les déviances observées sur la parole dysarthrique sont assez semblables à celles observées en parole spontanée, il est tout à fait envisageable de transposer ce système sur la parole spontanée en vue d'une recherche des zones de réduction. Enfin, l'amélioration des connaissances de ces phénomènes pourrait, à terme, permettre d'améliorer les outils d'alignement et de reconnaissance de ce type de parole atypique.

II.2 APPROCHE PSYCHOLINGUISTIQUE: LA RECONNAISSANCE DES MOTS

L'approche que nous allons aborder maintenant est tout autre. La psycholinguistique a pour objet d'étude les processus cognitifs mis en œuvre dans le traitement et la production du langage. Cet objectif est évidemment éminemment complexe car il demande de répondre à de nombreuses questions dont la rapidité du traitement de l'information, la séquentialité de ce traitement, la répartition et l'organisation des informations stockées, les processus d'accès au lexique, etc.

Pour répondre à ces questions la mise au point de méthodologies rigoureuses est nécessaire. Notamment, l'enregistrement des temps de réaction, permettant de comprendre le déroulement temporel du traitement des informations¹. Mais également, le contrôle des objets linguistiques manipulés est une condition nécessaire à l'interprétation des résultats. Tout facteur de variation étranger à la variable analysée doit être éliminé de façon à pouvoir interpréter correctement les effets observés.

Toutefois, ces contraintes peuvent jeter un voile sur la portée des résultats obtenus dans ces conditions. La statistique nous apprend que les effets observés sont très dépendants de l'échantillon recueilli et que, pour interpréter les résultats, l'échantillon doit être représentatif de la population. Il en va de même pour les expérimentations faites de façon très rigoureuse : les effets observés ne sont valables que pour le cadre expérimental dans lequel ils ont été observés. De quoi les stimuli expérimentaux sélectionnés sont-ils représentatifs? Sont-ils représentatifs des mots de langues? Sont-ils représentatifs des conditions dans lesquelles nous traitons la parole habituellement?

Les travaux en psycholinguistique ont apporté de nombreuses informations concernant les processus de traitement de la parole:

- nous sommes plus rapides pour produire un mot qu'un non-mot (Forster and Chambers, 1973, nous avons donc un lexique mental)
- nous sommes plus rapides pour reconnaître un mot fréquent qu'un mot rare (Forster and Chambers, 1973, les mots stockés sont donc organisés)
- nous sommes capables d'identifier un mot avant la fin de sa production (Modèle de la Cohorte, (Marslen-Wilson and Welsh, 1978))

¹ Toutefois assez imprécis malgré tout et sensible à de nombreux facteurs. La chronométrie mentale laisse actuellement place à des méthodologies plus fiables (EEG, mouvements oculaires, etc.).

- dans le processus de traitement, il y a des effets facilitateurs et inhibiteurs qui sont fonction de la proximité des mots (voisinage phonologique, Luce and Pisoni, 1998)
- etc.

Si la plupart de ces résultats sont assez fiables, d'autres semblent beaucoup moins clairs et les effets varient en fonction des expériences (cf. voisinage phonologique, etc.).

La question posée se reformule ainsi: l'effet d'un mécanisme mis en lumière dans des conditions expérimentales spécifiques est-il transposable à toutes les situations de parole? Y'a-t-il des effets robustes? Y'a-t-il une hiérarchie des processus de traitement ajustable par l'auditeur selon la situation de parole et la nature du substrat linguistique?

Contribution: Accès lexical: quelles unités?

Meunier C., Frauenfelder U., Content A., Kearns R. (1997) "The locus of the syllable effect: Prelexical or Lexical?", Proceedings of the Fifth European Conference on Speech Communication and Technology, Rhodes, Vol. 5, pp. 2851-2854.

Content, A., Meunier, C., Kearns, R., Frauenfelder, U.H. (2001) "Sequence detection in pseudowords in French: Where is the syllable effect?" Language and Cognitive Processes, 16, 5/6, 609-636.

La plupart des locuteurs d'une langue sont capables d'identifier les phonèmes présents au sein d'un mot. Toutefois, les recherches sur l'acquisition du langage semblent montrer que cette capacité ne serait pas effective avant l'âge de 6 ou 7 ans, soit l'âge correspondant approximativement à l'apprentissage de la lecture. En revanche, la capacité à extraire et à identifier les syllabes est, elle, beaucoup plus précoce et préservée chez les personnes illettrées (Morais et al., 1986). Il est donc possible que notre capacité à extraire et identifier les phonèmes soit une conséquence de l'apprentissage de la lecture, en particulier, des écritures alphabétiques (Labruno, 2005). L'utilisation du phonème dans le traitement perceptif pourrait donc être mise en question. L'unité syllabique pourrait s'avérer moins variable et présenter des caractéristiques plus robustes que celles du phonème.

De nombreux travaux, issus de la psycholinguistique, sont venus étayer cette hypothèse. Segui et al. (1981) montrent que les auditeurs sont plus rapides pour détecter le phonème /b/ en position initiale lorsqu'il se situe dans une syllabe simple comme /ba/, tandis qu'ils sont plus lents pour détecter ce même phonème dans des syllabes plus complexes (/bra/ ou /bal/). Ces résultats semblent suggérer que l'identification du phonème se fait *après* celle de la syllabe. Par ailleurs, si l'identification de voyelles extraites d'un corpus de parole spontanée est parfois très mauvaise (20% d'identification pour la voyelle [a]), elle est bien meilleure (75%) lorsque l'on y ajoute le contexte syllabique (Meunier & Floccia, 1999). Une partie importante de la variabilité serait ainsi due à la coarticulation au sein de la syllabe.

A l'aide d'un paradigme astucieux, l'équipe de Jacques Mehler a pu confirmer cette hypothèse (Mehler et al., 1981). L'expérience comprenait des cibles syllabiques de type CV ou CVC (PA ou PAL) ainsi que des mots porteurs bisyllabiques dont la première syllabe était elle-même de type CV ou CVC (*palace* PA.LACE; *palmier* PAL.MIER). Ainsi les deux cibles sont présentes dans les deux mots porteurs, mais dans un cas la cible est conforme à la structure syllabique de la première syllabe du mot porteur (PA dans PA.LACE et PAL dans PAL.MIER) alors que dans l'autre cas, la cible, bien que présente, n'est pas conforme à la structure syllabique de la première syllabe du mot porteur (PA dans PAL.MIER et PAL dans PA.LACE) (figure 2.h).

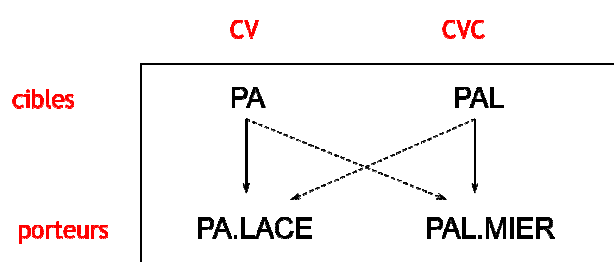


Figure 2.h: paradigme de l'expérience PALACE-PALMIER

La tâche des sujets était donc de détecter le plus rapidement possible une séquence cible (ex: /pa/ ou /pal/) dans un mot porteur prononcé (ex: palace ou palmier).

L'hypothèse était la suivante: si les sujets utilisent la syllabe comme unité permettant d'accéder au lexique, la similitude de la structure syllabique entre cible et porteur devrait faciliter la détection. Autrement dit, les temps de réaction devraient être plus rapides lorsque la cible et le porteur ont la même structure syllabique. En revanche, si les sujets font une détection purement phonémique, les temps de réaction devraient être plus rapide pour la cible PA, quelle que soit le mot porteur dans la mesure où le nombre de phonèmes à détecter est moindre.

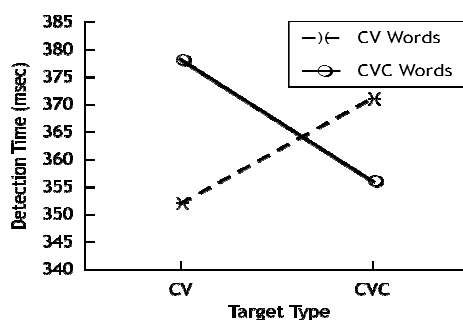


Figure 2.i: effet syllabique (Mehler, Dommergues, Frauenfelder & Segui, 1981)

Les résultats obtenus par l'équipe de Mehler confirment l'hypothèse syllabique (figure 2.i). Il a donc été conclu, à cette époque, que la syllabe constituerait une unité plus robuste que le

phonème pour l'accès au lexique. Ce résultat suppose également que l'identification des phonèmes serait post-syllabique.

Ces résultats ont été répliqués et l'effet semble probant pour de nombreuses langues pour lesquelles la frontière syllabique n'est pas ambiguë (Espagnol, Catalan, Portugais, Hollandais). A l'inverse, l'effet syllabique n'est pas retrouvé dans des langues pour lesquelles la frontière syllabique est ambiguë (anglais, japonais).

En 1994, avec l'équipe d'U. Frauenfelder et A. Content, nous avons tenté de répliquer l'expérience de 1981 en remplaçant les mots porteurs par des pseudo-mots. Il était question ici de vérifier si l'effet syllabique était pré- ou post-lexical. Si nous obtenions le même effet qu'en 1981 avec des pseudo-mots, cela signifiait que le traitement était bien pré-lexical et que la syllabe était bien une unité permettant l'accès au lexique. Dans le cas inverse, il fallait admettre que l'effet syllabique n'était peut-être dû qu'à un traitement post-lexical.

Nous avons par ailleurs apporté d'autres modifications par rapport à l'expérience de 1981: nous avons fait varier la consonne pivotale des mots porteurs (dans l'expérience de 1981, toutes les consonnes pivotales étaient des liquides, /r/ ou /l/). Nous avons également fait varier la voyelle de la cible et des porteurs (en 1981, seule la voyelle /a/ était utilisée). Enfin, nous avons introduit des distracteurs (expérience E2 à E4) de façon à ce que les sujets n'utilisent pas de stratégies spécifiques pour l'exécution de la tâche.

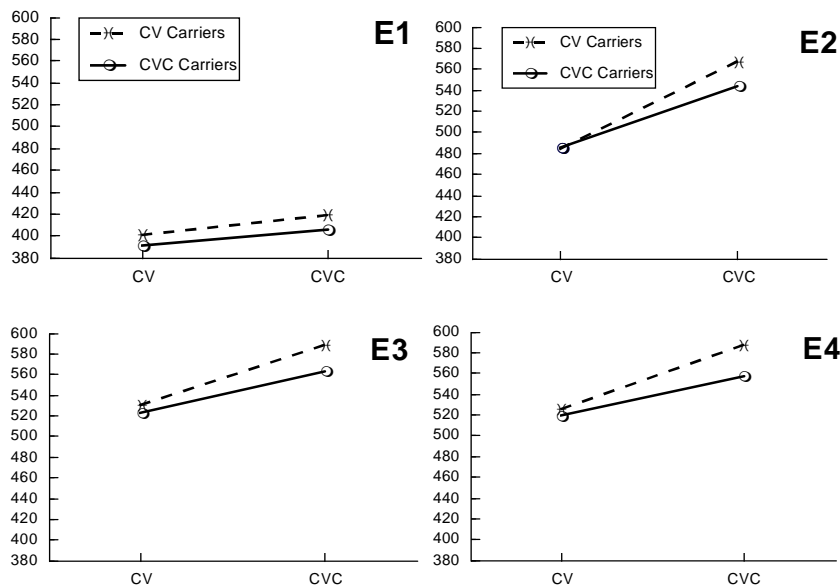


Figure 2.j: Temps de réaction recueillis pour les quatre expériences.

Nos résultats ne montrent aucun effet syllabique (figure 2.j), mais plutôt un effet de cible (la cible CV est toujours détectée plus rapidement). Nous relevons également un effet du porteur: les cibles sont détectées plus rapidement dans un porteur CVC (dont les phonèmes sont, en

général plus courts). L'ensemble de ces résultats évoquent un **traitement de nature phonémique**. L'expérience E1, réalisées sans distracteurs, montre des temps de réaction beaucoup plus rapides, mais pas d'effet syllabique. Il en va de même pour E2, avec les distracteurs. Dans les expériences E3 et E4, la présentation des stimuli a été bloquée respectivement par consonne pivotale (E3) et par voyelle (E4). Dans aucune de ces conditions nous n'avons pu répliquer l'effet syllabique.

Nous avons regardé plus en détail les résultats de l'expérience E3 (figure 2.k). Si l'on distingue les résultats obtenus pour chacun des types de consonne pivotale (liquide, fricative et occlusive) d'une part, et les sujets rapides des sujets lents d'autre part, il apparaît qu'un effet syllabique ressort lorsque la consonne pivotale est une liquide. Cet effet est plus fort chez les sujets lents. Chez ses même sujets lents, l'effet syllabique apparaît également légèrement pour les occlusives. L'effet ne ressort jamais pour les fricatives.

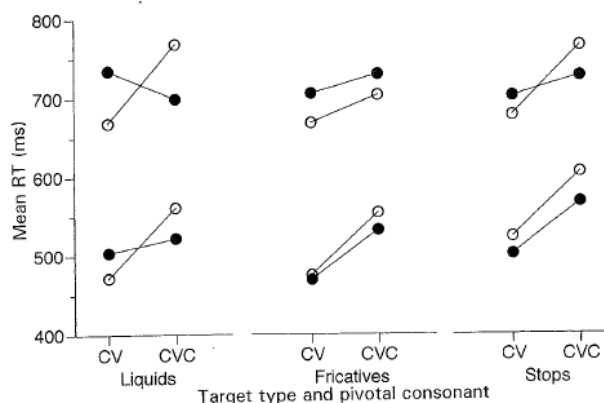


Figure 2.k: Temps de réaction pour la tâche de détection de cible dans l'expérience E3 (blocage par consonne pivotale). Distinction entre sujets lents (en haut) et sujets rapides (en bas).

L'interprétation de ces résultats reste très complexe. Si l'on tient compte des résultats obtenus en 1981, il semble que l'effet syllabique soit limité à des mots dont la consonne pivotale est une liquide (et encore, même dans ce cas l'effet reste fragile). Par ailleurs, seuls les sujets lents semblent montrer cet effet, ce qui suggérerait que l'effet n'est pas précoce. Ces résultats suggèrent un processus perceptif de type mixte (syllabique/phonémique) et rejette l'hypothèse d'un mécanisme de classification purement syllabique. Deux explications sont possibles concernant l'effet syllabique. La première, soutenue par Mehler et al. (1981) propose que la détection de la syllabe intervienne à un stade très précoce du traitement. Une version plus nuancée de cette hypothèse est proposée par Sebastián-Gallés et al. (1992) et Dupoux (1993). Elle suggère que le processus de perception est basé sur de petites unités et que les auditeurs peuvent évaluer en permanence l'unité la plus adaptée avec l'information acoustico-phonétique reçue. Nos résultats vont dans le sens de cette hypothèse.

II.3 BILAN

L'ensemble de mes contributions concernant ces deux aspects méthodologiques liés à l'étude de la parole sont à la fois bien distinctes et complémentaires.

Le travail d'expertise et d'annotation des corpus permet de mieux comprendre les réalisations effectives de la parole dans différentes situations. Cette expérience est à l'origine de ma conception de la variation de la parole comme paramètre d'ajustement et d'adaptation du locuteur à une situation donnée. Elle m'a permis d'enrichir mes connaissances sur le signal de parole et de développer un guide permettant l'annotation phonétique manuelle. Cette approche est indispensable pour appréhender les phénomènes spécifiques de la production de la parole sous tous ses aspects et ainsi décrire le processus d'adaptation des locuteurs dans cette tâche. Toutefois, comme nous l'avions mentionné, cette approche ne permet pas de comprendre le processus de compréhension dans toute sa complexité lors du traitement de la parole spontanée.

Mon séjour à l'Université de Genève en collaboration avec l'équipe de psycholinguistes m'a ouvert également des perspectives et une méthodologie nouvelle. La connaissance des modèles de perception de la parole m'a ainsi permis de mieux intégrer les questions complexes liées au traitement de l'information linguistique. Il m'est ainsi apparu que les processus de traitement sont très délicats à cerner et que les résultats obtenus par le biais des expérimentations sont extrêmement sensibles au matériel linguistique utilisé, mais aussi à la tâche qui est demandé aux sujets des expériences. Il apparaît que souvent, la réponse donnée par le sujet est plus déterminée par la tâche que par l'objet linguistique à traiter.

De nombreux travaux tentent désormais de concilier ces deux approches. C'est dans cette perspective que s'est inscrite ma collaboration avec Yao Yao (Hong Kong Polytechnique University) dans le cadre d'une recherche sur l'effet de la densité du voisinage phonologique dans la production des mots. Cet effet est mesuré sur la réduction vocalique dans des productions spontanée en français (corpus CID). Dans cette optique, j'envisage donc de développer des travaux permettant l'étude expérimentale de ces aspects (voir *Perspectives* à la fin de ce document).

III. PARLER: LES SONS DE LA PAROLE

"Dès que je pense le monde, je l'organise et je ne peux pas penser le monde sans classer les choses ou les phénomènes qui, à mon sens, le composent." Georges Vignaux, Le démon du classement.

Parler est une activité proprement humaine qui nécessite des dispositions physiologiques et des compétences spécifiques relatives à un système abstrait. Le langage est ainsi une construction complexe faisant interagir les caractéristiques du vivant et celles des représentations abstraites. Les propriétés qui régissent le fonctionnement et la construction des langues sont de nature abstraite (propriétés syntaxiques, phonologiques, lexicales, etc.) mais leur actualisation est réalisée par le corps et dépend donc de ses limites et de ses contraintes. Longtemps la linguistique a séparé ces deux aspects, Saussure (1916) sous la distinction langue/parole et Chomsky (1957) sous celle de compétence/performance. Dans cette approche, la parole n'est alors qu'une actualisation de la langue et il n'est pas admis que des propriétés physiques et physiologiques puissent avoir des conséquences sur la structure de la langue. Pourtant, on sait que la structure des systèmes vocaliques est conditionnée par les limites physiologiques du tractus vocal (il est possible de produire des /a/ plus ou moins ouverts, mais il y a une limite à l'ouverture de la mandibule).

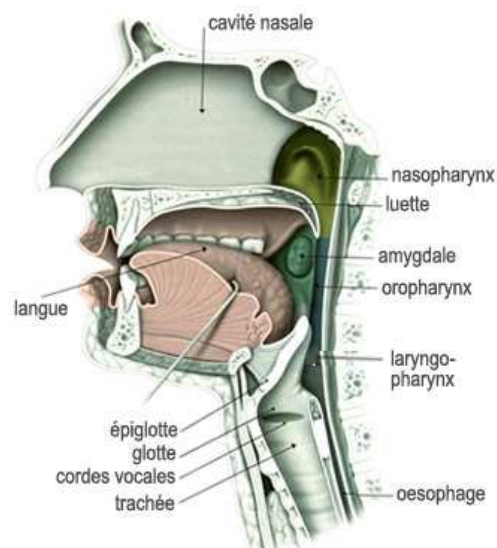


FIGURE 3.a – Tractus vocal de l'humain. Ensemble des organes servant, entre autres, à la production de la parole (<http://lecerveau.mcgill.ca>)

Le corps humain dispose donc d'un ensemble *d'outils* (figure 3.a) permettant la production articulée d'un langage d'une grande complexité. L'ensemble des organes utilisés pour la production de la parole ont probablement évolué à partir de structures initiales simples (Gould & Vrba, 1982¹).

Les cordes vocales permettent en premier lieu d'obturer l'entrée des poumons; le voile du palais conditionne la respiration nasale ou buccale; la mandibule, la langue, les dents sont d'abord utiles pour la mastication, etc. Ainsi, la plupart des organes indispensables à la production de la parole, ont en premier lieu des fonctions indispensables à la survie de l'être humain: respirer, manger. La parole est donc une sorte *d'activité parasite* venant se greffer sur les fonctions biologiques du corps humain. Mais cette activité parasite est tellement ancienne qu'elle semble désormais totalement intégrée dans les fonctions biologiques fondamentales de l'humain. Ainsi, l'acquisition du langage est aussi irrépressible que l'acquisition de la marche, pour peu que l'environnement linguistique permette son développement (Pinker, 1994).

Les organes du tractus vocal, communs à tous les locuteurs quelle que soit leur langue, permettent de représenter un système de sons organisés dans lequel chaque langue puise son propre inventaire. La structure de l'inventaire des sons dans une langue n'est pas aléatoire mais suit une logique universelle relative aux contraintes articulatoires et à la distinctivité nécessaire au sein de chaque système (Johnson, 2000). Alors que les systèmes vocaliques peuvent contenir jusqu'à 28 voyelles, les deux tiers des langues n'en font apparaître que 5 ou 7 ((Vallée et al., 1999). Si un système ne présente que trois voyelles, ces voyelles occuperont les positions articulatoires les plus éloignées (Lindblom, 1986; Schwartz et al., 1997). De même, on sait que les systèmes incluent massivement les occlusives dans leur inventaire au détriment des autres classes de consonnes (Vallée et al., 1999).

Ainsi, les systèmes de sons des langues relèvent de propriétés abstraites et de contraintes propres à la viabilité du système. Mais ils relèvent également de propriétés et de contraintes physiologiques. Par ailleurs, à l'intérieur de chaque langue, la production des sons est soumise à des contraintes dues 1/ au fait que les sons sont produits conjointement (coarticulation) et 2/ aux conditions dans lesquelles le locuteur parle (débit, caractère naturel, etc.). Des travaux

¹ Stephen J. Gould et Elizabeth Vrba, expliquent comment, au cours de l'évolution des espèces, des caractéristiques physiques complexes peuvent apparaître et évoluer à partir de structures initiales simples.

récents ont également montré que le développement et les caractéristiques des systèmes phonologiques pouvaient être dépendants du milieu environnant (Maddieson et al., 2013).

Pour ce qui concerne la production des sons du langage, il y a donc une interaction certaine entre un système aux propriétés abstraites et des contraintes physiologiques et environnementales. C'est dans cet intervalle que je propose d'exposer quelques travaux dont l'objet est ici de décrire l'usage et les formes de réalisation des sons du français.

III.1 LES SONS ET LEUR USAGE

Indépendamment des aspects strictement phonétiques (au sens de la description physique des sons), la façon dont les sons d'une langue sont utilisés dans le langage oral joue probablement un rôle important pour leur changement, leur évolution ou encore leur variation. En effet, les gestes articulatoires réalisés lors de l'émission de la parole sont répétés durant toute notre vie avec une fréquence peu quantifiable. Ces gestes sont conditionnés par la fréquence des sons, leur combinatoire et donc leur enchaînement. Ces paramètres étant spécifiques aux langues il est difficile d'imaginer qu'ils n'aient pas une incidence sur notre production. Chaque langue développerait alors des routines articulatoires qui conditionneraient la production. Ces routines pourraient donc être d'une grande importance dans toutes les situations où le locuteur doit s'adapter (apprentissage d'une langue, situations de communication, etc.).

Dans les travaux mentionnés ci-dessous, nous nous sommes intéressés à la fréquence de différents types d'occurrences dans le CID.

Contributions:	Occurrences en parole spontanée
Meunier, C. (2012) " Contexte et nature des réalisations phonétiques en parole conversationnelle ", Actes des Journées d'Etude sur la Parole, Grenoble (France), juin 2012, 1-8.	

Dans cette perspective, nous avons réalisé une étude des occurrences de mots, syllabes ou encore phonèmes dans la parole conversationnelle en français (corpus CID). Les 8 heures du CID contiennent 6 611 formes lexicales¹ différentes tandis que la totalité des occurrences¹ est de

¹ Par "formes lexicales", nous entendons l'ensemble des types de mots (au sens de *Type*), indépendamment du nombre d'apparitions dans le corpus.

102 457. La moitié des occurrences lexicales du corpus totalise seulement 39 formes différentes (soit, 0,6% des formes lexicales du corpus). La distribution des formes lexicales présentes dans le CID est semblable à celle que l'on trouve dans l'ensemble des corpus oraux spontanés et des grands corpus textuels (Lexique, New et al., 2001). Les phonèmes les plus fréquents sont les mêmes (/r a ε l s i t d e/...), parfois dans des ordres différents, mais les grandes tendances sont préservées quel que soit le type de corpus. On y retrouve les caractéristiques de la loi de Zipf avec des mots dont la fréquence est inversement proportionnelle à leur rang dans le corpus. Les formes lexicales n'apparaissant qu'une seule fois dans le corpus sont au nombre de 3 259 tandis que la forme la plus fréquente apparaît 3 130 fois, ce qui est comparable à d'autres corpus de parole spontanée (Torreira et al., 2010). En revanche, la spécificité des mots fréquents des corpus conversationnels repose sur la nature de ces mots (*ouais, je, tu*, etc.) que l'on ne retrouve pas avec cette fréquence dans des corpus radiophoniques ou des textes journalistiques.

Forme	est	c'	ouais	et	de	tu	pas	je	ça	le
Nbre	3130	3018	2916	2679	2033	2027	1895	1893	1817	1655
% d'occurrences	3,05	2,95	2,85	2,61	1,98	1,98	1,85	1,85	1,77	1,62

TABLE 3.a: Occurrences des 10 formes lexicales les plus fréquentes

On remarquera que les mots les plus fréquents (table 3.a) sont courts et pour la plupart des mots fonction (déterminants, conjonction, etc.). Plusieurs travaux ont pu montrer que ces caractéristiques lexicales ont une influence sur les productions phonétiques (Johnson, 2004) (Meunier and Espesser, 2011a). Sur les 39 formes les plus fréquentes, seulement 3 sont bisyllabiques (*était, enfin, avait*) et aucune n'est un nom. 57% des mots du corpus sont des monosyllabes. Les quatre compositions syllabiques les plus fréquentes totalisent plus de 50% du corpus et regroupent des mono- et des bisyllabes (table 3.b).

	monosyllabes		bisyllabes	
Forme	CV	V	CV.CV	V.CV
%	22	17,5	7	6

TABLE 3.b: Décomposition syllabique des mots les plus fréquents du corpus.

¹ Par "occurrences", nous entendons le nombre d'apparitions des mots dans le corpus (comptage au sens de *Token*).

Ces caractéristiques rendent l'exploitation statistique des données très délicate. En effet, plusieurs facteurs sont en interaction et il est difficile de les exploiter séparément. Comparer les mots de contenu fréquents et rares revient à comparer très peu de mots répétés de très nombreuses fois à une grande quantité de mots répétés très peu de fois.

La décomposition syllabique du corpus montre un total de 139 751 syllabes¹. La fréquence syllabique est, là encore, conforme à ce que l'on peut trouver dans des corpus de textes journalistiques ou dans les bases de données lexicales (Goldman et al., 1996), la structure CV étant de loin la plus fréquente et représentant, à elle seule, plus de la moitié des syllabes produites. Les six structures syllabiques les plus fréquentes représentent 99% des syllabes du corpus, ce qui rend les autres structures très marginales (table 3.c).

Forme	CV	V	CVC	CCV	CCVC	VC
%	60,5	13	11,5	10,5	2	1,5

TABLE 3.c: Structures syllabiques les plus fréquentes.

Les phonèmes sont issus de la transcription du corpus pour laquelle les experts avaient la possibilité d'indiquer les élisions (grâce à la TOE, Bertrand et al., 2008). Les 272 166 phonèmes (53% de consonnes) sont donc ceux qui ont été perçus et transcrits par les experts. Pour le CID, les voyelles à timbre variable ont été regroupées². La fréquence des phonèmes produits dans le CID est sensiblement comparable à celle que l'on peut trouver dans différentes bases de données (ici *Lexique 2*³, figure 3.b).

La voyelle /e/ est surreprésentée dans le CID en raison de la fréquence des mots *est* (sous la forme *c'est*, table 3.a) et *ouais* dans ce corpus alors que ce dernier mot est évidemment absent des corpus utilisés par *Lexique*. En revanche, /r/ est sous-représenté dans le CID. Cela peut-être dû au fait que cette consonne fait partie des élisions les plus fréquentes (voir plus loin). On notera également la forte représentation de /w/, ce qui, là encore, peut-être dû à la fréquence du mot *ouais*.

1 Syllabation du corpus effectuée à l'aide du syllabeur développé au LPL par B. Bigi (Bigi et al., 2010). Cette syllabation est indépendante des frontières lexicales.

2 /e ε/, /o o/, /ø œ ə/, /a a/. En vue d'une comparaison, les mêmes regroupements ont été effectués pour les fréquences de *Lexique*.

³ *Lexique 2*, site réalisé par Boris New & Christophe Pallier (<http://www.lexique.org>)

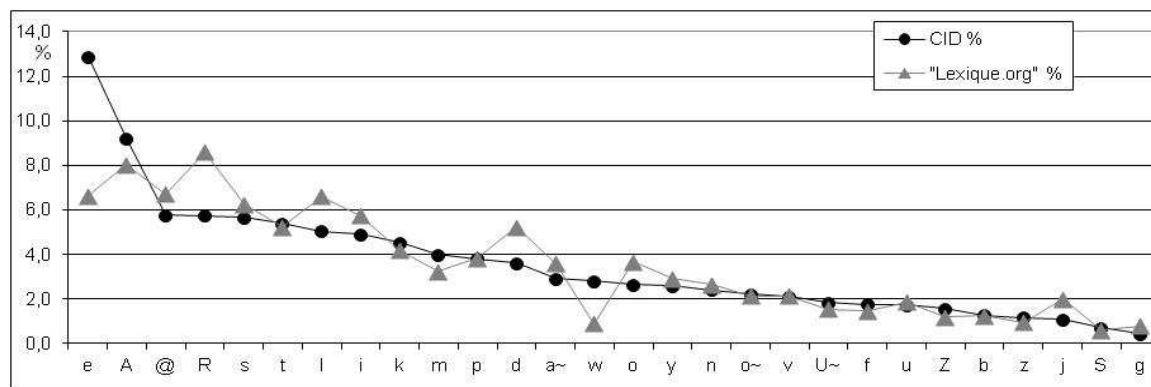


FIGURE 3.b: Fréquences des phonèmes du CID comparées à celles extraites de la base *Lexique* (New et al., 2001)(codé en SAMPA¹).

La majorité des phonèmes sont réalisés dans un nombre de mots très restreints. Par exemple, 50% des réalisations de la voyelle /a/ se trouvent dans seulement 13 mots différents (figure 3.c) et 20% des réalisations de la voyelle /y/ se trouvent dans le pronom *tu* (très fréquent en parole conversationnel). Là encore, ces mots sont essentiellement des mots fonction monosyllabiques. Le support lexical de ce type de corpus est ainsi très éloigné de celui utilisé dans des corpus construits.

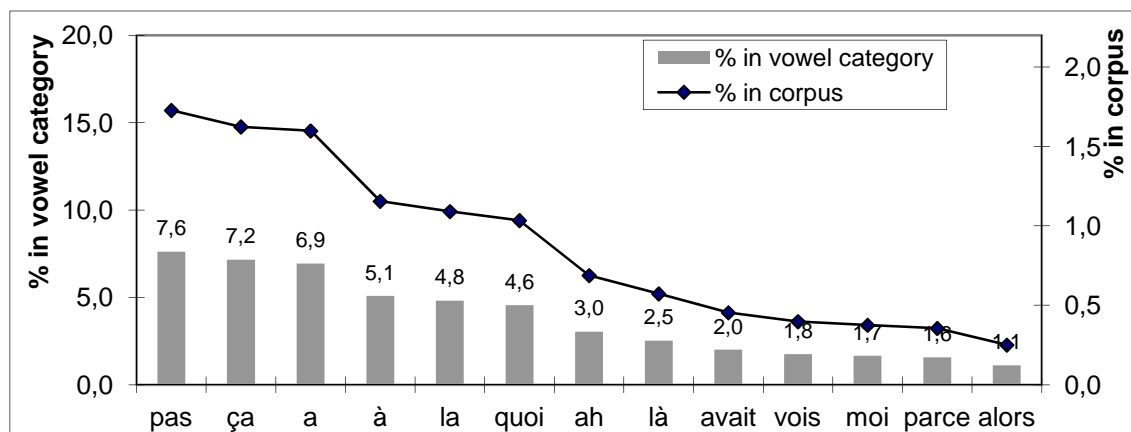


Figure 3.c: mots dans lesquels on trouve 50% des réalisations de la voyelle /a/ pour tout le corpus.

Les experts transcripateurs du CID avaient la possibilité de coder les réalisations particulières (variantes perçues) ainsi que les élisions identifiées. 2 948 réalisations particulières ont été

¹ Correspondance SAMPA-IPA:

Voyelles	IPA	i	y	u	e	ø	o	ɛ	œ	ɔ	a	ɑ	ə	ã	ẽ	æ̃	õ			
	SAMPA	i	y	u	e	2	o	E	9	O	a	A	@	a~	e~	9~	o~			
Consonnes	IPA	p	t	k	b	d	g	f	s	ʃ	v	z	ʒ	m	n	l	r	j	w	ɥ
	SAMPA	p	t	k	b	d	g	f	s	S	v	z	Z	m	n	l	R	j	w	H

codées. 20% de ces réalisations concernent les cas spécifiques du pronom *je* suivi d'une consonne non voisée (/s/ la plupart du temps). Il s'agit donc pour l'essentiel des séquences *je sais* (produit /ʃe/) ou *je suis* (produit /ʃti/). Les deux phonèmes sont ici fusionnés (lieu d'articulation du premier phonème /ʒ/ et voisement de la deuxième consonne /s/, le tout donnant un seul phonème /ʃ/. Cette forme, très fréquente, est particulièrement bien identifiée par les auditeurs et peut être considérée comme une forme figée, assez prédictible. Les autres réalisations codées sont en lien avec l'accent des locuteurs. Notamment, de nombreux /ə/ ont été ajoutés comme réalisation particulière là où le transcripateur s'attendait à son absence. De même le pronom *tu* est souvent produit /ti/ ce qui est une spécificité des locuteurs du sud-est (Jamin et al., 2006).

10925 élisions ont été codées. Les dix élisions les plus fréquentes (table 3.d) représentent 99% des élisions codées dans le corpus. /ə/ est clairement le phonème le plus souvent identifié comme manquant, ce qui n'est pas surprenant puisque ce symbole comprend le schwa dont la présence ou l'absence sont souvent bien identifiées par les transcripateurs.

phonème	@	l	y	R	a~	v	e	i	u	d
%	35,8	19,1	8,4	8,6	5	3,1	3,1	2,1	13	1,1

TABLE 3.d: Les dix élisions les plus fréquemment codées par les experts

Ces résultats sont conformes à ce que l'on trouve dans la plupart des corpus de parole continue. On note quelques particularités dues au style conversationnel comme la sur-représentation de la voyelle /e/. Notons toutefois que la très grande majorité des formes et des occurrences lexicales que l'on trouve dans ce type de corpus (mots fonction monosyllabiques) sont, la plupart du temps absents des travaux phonétiques ou psycholinguistiques portant sur des corpus construits. Notons également que, dans une conversation, la majorité de ces mots ne porte pas l'information principale du message linguistique.

Contributions: **Occurrence des gestes articulatoires en parole spontanée**

Meunier, C., & Espesser, R. (2011) "Is phoneme inventory a good predictor for vocal tract use in casual speech?", Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011), Hong-Kong, China, August 2011, 1370-1373.

Dans ce travail, je me suis plus précisément intéressée aux occurrences des gestes articulatoires produits dans le même corpus de parole conversationnelle (CID). En effet, si l'inventaire des phonèmes d'une langue semble être régi par des contraintes phonologiques et diachroniques,

on peut également faire l'hypothèse que cet inventaire, et surtout les gestes articulatoires représentatifs (Browman and Goldstein, 1992), est dépendant de l'usage des phonèmes dans la langue. Nous avons donc cherché à comparer: 1/ la représentation des gestes articulatoires répertoriés dans l'inventaire phonologique du français avec 2/ la proportion de ces mêmes gestes dans le lexique du français et dans un corpus de parole spontanée (Meunier and Espesser, 2011). Chaque langue diffère selon les spécificités de son inventaire phonologique. Par exemple, certains dialectes arabes sont caractérisés par des positions articulatoires très postérieures (uvulaires, pharyngales, glottales), tandis que le suédois possède de nombreuses voyelles antérieures. En français il n'y a pas de consonne très postérieure (excepté /ʀ/, uvulaire) et un tiers des voyelles sont antérieures tandis qu'un autre tiers est arrondi. Ces particularités donnent une représentation globale de l'utilisation du tractus vocal dans ces langues. Néanmoins, cet inventaire donne-t-il une représentation fidèle des positions articulatoires lors de la production de parole spontanée? L'inventaire des phonèmes dans une langue attribue le même poids à chaque propriété de phonème alors que les phonèmes sont diversement présents dans les mots et les mots diversement présents dans le discours. Dans quelle mesure, les positions articulatoires présentes dans l'inventaire des phonèmes sont-elles modifiées par la fréquence des phonèmes dans les mots et la fréquence des mots dans le discours? L'étude précédente, portant sur l'occurrence des phonèmes et des mots en conversation, nous invite à penser que la fréquence de certains mots pourrait influencer l'utilisation de certains gestes articulatoires.

Pour répondre à ces questions, nous avons comparé les fréquences des positions articulatoires pour trois conditions différentes:

1. **PI**: Inventaire Phonologique du français (15 voyelles et 19 consonnes)
2. **WO**: base de données lexicales des mots du français (VanRullen et al., 2005)
3. **CO**: corpus de parole conversationnel, le CID (Bertrand et al., 2008).

Les positions articulatoires ont été déterminées d'après la terminologie de la Phonologie Articulatoire (AP , Browman and Goldstein, 1992). Les Gestes de l'AP se présentent comme une représentation abstraite de l'usage du tractus vocal durant la production de la parole. Nous n'avons pas utilisé les traits acoustiques ou articulatoires car ils sont peu précis et trop éloignés de la production de la parole. Par conséquent, ce travail ne fournit pas d'informations sur l'articulation *réelle* des sons du français mais sur les positions attendues ou du moins canoniques.

L'AP propose que la représentation phonologique des sons de la parole soit exprimée sous la forme d'un ensemble de variables et dimensions du tractus vocal (*tract variables and*

dimensions). Les dimensions spécifient la position des organes (ouvert, fermé, palatal, etc.). Les variables sont:

- Lip Protrusion (LP)
- Lip Aperture (LA)
- Tongue Tip Constriction Location (TTCL)
- Tongue Tip Constriction Degree (TTCD)
- Tongue Body Constriction Location (TBCL)
- Tongue Body Constriction Degree (TBCD)
- Velum (VEL)
- Glottis (GLO).

Pour cette étude, nous avons utilisé les variables et dimensions suivantes:

- VEL closed, wide
- GLO closed, wide
- TTCL alveolar, postalveolar
- TBCL postalveolar, palatal, velar, uvular
- LP protruded

Une analyse des phonèmes du français a permis de caractériser l'ensemble des variables et dimensions pour chaque phonème (exemple table 3.e).

	LP	TTCL	TBCL	VEL	GLO
n	-	alveolar	-	wide	closed
k	-	-	velar	closed	wide
b	-	-	-	closed	closed
ŋ	-	-	postalveolar	closed	wide
u	protruded	-	velar	closed	closed

Table 3.e: Quelques exemples de dimensions affectées aux phonèmes du français.

Ensuite, un coefficient de fréquence (en pourcentage) pour chaque dimension a été établi en fonction des trois conditions (système phonologique **Pi**, lexique **Wo** et corpus conversationnel **Co**). La fréquence de la dimension est calculée comme suit:

- (Nombre d'occurrence pour une dimension) / (nombre de phonèmes dans la condition) * 100
- avec Pi = 34 phonèmes, Wo = 339 813 phonèmes et Co = 301 918 phonèmes.

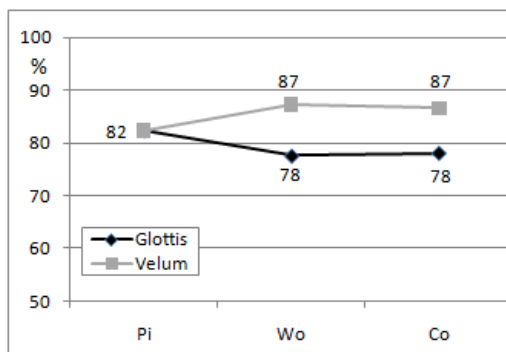


Figure 3.d frequency (in %) of close position for Glottis and Velum in the three datasets.

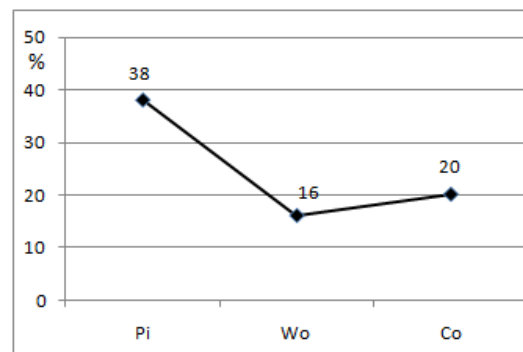


Figure 3.e : frequency (in %) of protruded position in the three datasets.

On constate que les différences observées sont relativement faibles pour la plupart des variables. Autrement dit, on peut concevoir que l'inventaire phonologique du français reflète assez bien la proportion des gestes qui sont utilisés dans la conversation. Ou, d'une certaine façon, les phonèmes présents en français sont utilisés de façon équilibrée dans le lexique et dans le discours.

Notons toutefois une augmentation de la position fermée de la glotte dans les mots et dans la conversation (figure 3.d) de même qu'une diminution de la position fermée du velum dans ces mêmes corpus. On observe, par ailleurs, que la position fermée de ces deux articulateurs est de loin la plus courante. Mais la modification la plus importante est celle du geste de protrusion (figure 3.e) la proportion de ce geste dans l'inventaire des phonèmes du français est nettement plus importante que celle relevée dans le lexique ou dans la conversation. Cette modification tient au fait que les voyelles les plus fréquentes en parole spontanée sont les voyelles d'avant non arrondies. Elles sont moins nombreuses dans l'inventaire, mais beaucoup plus fréquentes.

De façon surprenante, la proportion des gestes présents dans l'inventaire des phonèmes est peu affectée par la fréquence des phonèmes dans la langue et dans son usage courant. On peut ainsi supposer qu'il existe une sorte d'équilibre, non pas au niveau des phonèmes, mais au niveau des gestes présents dans une langue donnée. Si un phonème est très peu fréquent, les gestes qui le caractérisent sont sans doute représentés dans des phonèmes très fréquents. Ce travail doit être poursuivi dans différentes langues afin d'observer les spécificités de chaque langue, mais aussi une éventuelle tendance commune à l'utilisation des gestes. Autrement dit, des inventaires phonologiques variés pourraient tendre vers une utilisation comparable dans l'usage de la langue.

III.2 VARIATIONS ET REDUCTIONS EN PAROLE

À la lumière d'un siècle de travaux portant sur l'analyse de la parole, le constat est qu'il est impossible d'évoquer la réalisation des sons sans parler de leur variation. Nul ne le conteste, ni ne l'a, d'ailleurs, jamais contesté. Le débat concernant la variabilité de la parole ne porte donc pas sur son existence, mais sur son statut. Est-elle le résultat d'une actualisation physique, et donc non pertinente, des sons du langage ? Représente-t-elle du bruit sur un signal de parole comportant, par ailleurs, toute l'information nécessaire à la compréhension des messages ? Participe-t-elle pleinement, au contraire, au fonctionnement de la communication, à la structure et à l'évolution de la langue ? Est-elle gênante ou bien indispensable aux auditeurs pour

l'identification des sons de la chaîne parlée? La réponse à ces questions peut-être tour à tour positive ou négative tant la variation est multiforme.

Si la variabilité est partout dans la parole, la recherche *d'invariants*, autrement dit de caractéristiques stables et indépendantes de facteurs extrinsèques permettant de définir un phénomène, est une constante dans la démarche scientifique. Il semble d'ailleurs que la dualité « invariants - variabilité » ait suivi des évolutions théoriques et méthodologiques relativement semblables à l'intérieur des sciences humaines, sociales et cognitives au cours du XXème siècle (Lautrey et al., 2002). Autrement dit, la recherche d'invariants, dans un premier temps, puis un regain d'intérêt pour les phénomènes de variabilité ces dernières décennies.

Dans les premiers travaux de la linguistique moderne, Saussure (1916) distingue la *langue* de la *parole* : la langue est définie comme un code, tandis que la parole est l'utilisation de ce code. Cette distinction, fondement de la linguistique moderne, scelle la problématique générale de la relation existant entre la forme (représentations abstraites et conceptuelles) et la substance (matérialité et actualisation de ces représentations). La dualité invariants/variabilité peut alors être considérée comme une extension de la dichotomie langue/parole. La langue apparaît comme le domaine des invariants tandis que la parole serait le terrain de la variabilité (Fuchs, 2002) dans la mesure où la langue est un code partagé par tous les sujets d'une langue, tandis que la parole est l'expression (ou plutôt une infinité d'expressions possibles) de ce code.

En marge des travaux portant sur l'invariant, plusieurs axes de recherches ont traité le problème de la variabilité comme un phénomène interne et non externe à la langue. Dans un ouvrage qui fait référence concernant la question de la variabilité de la parole (Perkell & Klatt, 1986), Elman et McClelland (1986) soulignent qu'une partie importante des variations observées dans la parole ne sont pas dues au hasard. Ces auteurs s'intéressent particulièrement aux variations contextuelles qui, selon eux, sont fortement prédictibles et donc utiles aux auditeurs. La préoccupation est désormais de mettre en lumière les aspects de la variabilité qui jouent un rôle dans les mécanismes de perception de la parole.

Il y a en effet un lien historique et méthodologique qui a façonné notre façon de représenter les sons de la parole. Notamment, les caractéristiques des voyelles et des consonnes d'une langue sont toujours décrites comme des caractéristiques hors variation (donc issues de productions sans contexte). Il s'agit en fait de productions hyperarticulées (on parle aussi de "cibles articulatoires") produites dans un contexte qui impose cette hyperarticulation. Ces productions hyperarticulées sont ensuite considérées comme des prototypes à partir desquels les variations observées viennent modifier la nature phonétique des sons.

Or cette vision conditionne la façon dont nous appréhendons nos corpus et les productions orales. Nombreux sont ceux qui considèrent que les variations, quelles qu'elles soient, sont du bruit par rapport à des prototypes phonétiques. Nous sommes également nombreux à penser qu'on s'éloigne d'un contexte "normal" de production de la parole lorsque l'on est confronté à des réductions extrêmes produites dans un contexte conversationnel. Or cette réaction reflète souvent notre désarroi face à des réalisations que nous avons du mal 1/ à analyser et 2/ à interpréter.

Je distinguerai ci-dessous deux formes de variations de façon assez arbitraire: l'assimilation et la coarticulation, d'une part, et le phénomène de réduction, d'autre part. Cette distinction est conforme à la fois à la chronologie des travaux sur la variation, à leur forme respective, mais aussi au contexte qui favorise leur apparition.

III.2.1 Coarticulation et assimilation

Parmi les différents facteurs qui occasionnent la variabilité des sons de la parole, celui qui a probablement donné lieu au plus grand nombre d'études au cours du XXème siècle est le phénomène de coarticulation. Elle peut être décrite comme le chevauchement et l'interaction des différents articulateurs au cours de la production de segments phonétiques successifs (Farnetani, 1997). Il en résulte que la configuration du conduit vocal à un instant donné peut être mise en relation avec les caractéristiques de différents segments phonétiques. À l'inverse, chaque unité phonétique peut être matérialisée par des indices acoustiques répartis en différents points du signal acoustique (Nguyen, 2001). Par exemple, lors de la réalisation du mot sous /su/, la configuration du conduit vocal présente, dès le début de la fricative, les caractéristiques du trait d'arrondissement de la voyelle /u/ dont l'articulation est *anticipée*. En revanche, c'est par *persistance* que le trait sourd de l'occlusive /t/ se propage durant la réalisation du /t̚/ dans le mot /t̚ro/.

Un large débat a longtemps animé la communauté scientifique à propos de la distinction entre *coarticulation* et *assimilation*. Dans les deux cas, il s'agit de l'influence d'un son sur un autre. La coarticulation serait un phénomène dû à des contraintes physiologiques (synergie des articulateurs), et donc plutôt universel. En revanche, l'assimilation serait un phénomène de propagation des traits phonétiques selon les règles phonologiques présentes dans une langue donnée: le phénomène d'harmonie vocalique se définit comme la propagation des traits caractéristiques d'une voyelle à une voyelle avoisinante (Tranel, 1987). Il semble que ce phénomène soit spécifique à certaines langues. Le débat reste ouvert car certains travaux ont pu

montrer que les phénomènes de coarticulation n'étaient pas identiques dans toutes les langues et pouvaient être dépendants des systèmes phonologiques (Lubker and Gay, 1982).

Notons que l'assimilation endosse souvent un caractère plus linguistique ce qui est moins le cas pour la notion de coarticulation. De ce fait, on considère souvent l'assimilation comme un phénomène catégoriel, le /b/ prenant l'intégralité du trait non voisé dans le mot *obtus* /*optu*/. Or, on sait désormais que très souvent cette assimilation est graduelle et peut affecter la totalité comme une partie de la consonne.

L'assimilation et la coarticulation peuvent être considérées comme des variantes combinatoires. Les variantes combinatoires sont des variations dues au contexte phonétique mais ne présentant pas de caractère distinctif. Par exemple, la réalisation phonétique de la consonne [r] va dépendre du caractère sonore ou sourd de la consonne qui précède : le [r] sera réalisé sonore en contexte sonore [dʁɛ̃] et sourd en contexte sourd [tʁɛ̃]. Ces variations se distinguent des traits distinctifs: le trait de sonorité sera distinctif dans les mots « doux » (/du/) et « toux » (/tu/) car la permutation de ce trait entraîne un changement de sens (Durand, 2005). Les variantes libres sont des variations qui ne sont pas directement imputables à un facteur spécifique. Pour l'école fonctionnaliste (Troubetzkoy, 1939), l'ensemble de ces variantes n'a pas de fonction particulière dans la communication. Elles ne représentent donc pas un objet d'étude pour le linguiste à l'inverse des traits pertinents dont la fonction permet de distinguer les phonèmes et donc les unités lexicales d'une langue.

Nous verrons ci-dessous que nos travaux remettent en question le rôle perceptif des variations allophoniques.

Contribution: Assimilation et allophonie

Meunier C. (1999) "Recovering missing phonetic information from allophonic variations.", Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco.
Meunier C. (1997) "Voicing assimilation as a cue for cluster identification", Proceedings of the Fifth European Conference on Speech Communication and Technology, Rhodes, CDROM.

A la suite de mon doctorat, j'ai voulu tester la validité perceptive de certains phénomènes d'assimilation observés au sein des groupes de consonnes en français. Mon hypothèse était que plus un phénomène d'assimilation est fréquent, plus il a de poids dans le traitement perceptif des séquences de phonèmes.

J'avais pu observer que les assimilations de voisement classiquement décrites en français concernaient les groupes d'occlusives et de fricatives (donc des macro-classes phonétiques présentant une opposition de voisement). Or, au sein de ces groupes de consonnes, même si des

tendances nettes sont observées, l'assimilation n'est pas systématique: si le mot "obtus" est le plus souvent réalisé /optɥ/, le dévoisement du /b/ n'est pas systématique et l'assimilation peut également être partielle (si assimilation il y a). En revanche, dans des groupes du type occlusive+liquide (*train*) ou fricative+liquide (*frais*) l'assimilation est systématique: le /r/ est toujours dévoisé¹. J'ai donc fait l'hypothèse que cette assimilation était importante pour l'identification du groupe de consonnes. En effet, la systématisme d'une variante au sein d'une séquence de phonèmes est probablement intégrée par les auditeurs comme un trait caractéristique de cette séquence, comme ont pu le montrer certains travaux (Bradlow, 2002).

J'ai donc constitué 16 paires de stimuli comportant des groupes de consonnes du type occlusive+liquide avec une opposition de voisement sur l'occlusive du type:

GLACE /glas/ - **CLASSE** /klas/

DRAME /dram/ - **TRAME** /tram/

Dans cette première version (V1) enregistrée par un locuteur le voisement entre l'occlusive et la liquide est congruent: la liquide est voisée au contact de l'occlusive voisée et dévoisée en contexte dévoisé.

Une deuxième version (V2) a donc été créée en manipulant V1 (par cross-splicing): les liquides voisés et dévoisés ont été extraites de leur contexte et interverties, ce qui donne des stimuli pour lesquels l'information phonétique est incongruente (table 3.f).

Voiced stop		Unvoiced stop	
V1	V2'	V1	V2'
glace	gl(nvs)ace	classe	cl(vs)asse
g _v las	g _v las	k _v las	k _v las

Table 3.f: versions originale et manipulée des stimuli en fonction du trait de voisement de l'occlusive et de la réalisation allophonique de la liquide

Les résultats montrent que, indépendamment du voisement de l'occlusive, qui reste identique dans les deux versions, c'est le trait de voisement de la liquide qui semble déterminer l'identification du groupe de consonne comme voisé ou non voisé (figure 3.f).

Notons que l'effet de l'incongruence est plus fort lorsque le stimulus est de type Occ_nvs+Liq_vs. Cet effet est probablement dû au fait que les indices acoustiques représentant une occlusive non voisée en position initiale sont extrêmement réduits (juste un bruit d'explosion d'une dizaine de

¹ Notons toutefois que l'assimilation est plus prégnante pour /r/ que pour /l/. Le /l/ est également dévoisé mais plus souvent partiellement en fonction du contexte (consonantique et vocalique).

millisecondes), tandis qu'une occlusive voisée présente beaucoup plus d'informations acoustiques (tenue voisée + bruit d'explosion). Les indices acoustiques caractéristiques de la liquide ont alors un poids très important lorsque l'occlusive est non voisée. Cet effet est conforté par le fait que, très souvent en initial de mot, la tenue voisée d'une occlusive est réduite.

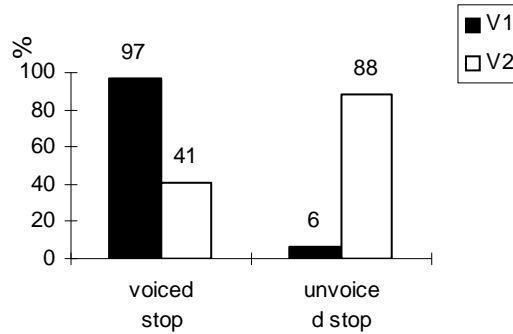


Figure 3.f: Word identification (in percentage) according to the stop voicing cue and the version type (V1 and V2). Only the "voiced" responses are reported here (each time a subject considered that the word started with a voiced stop).

Ainsi, la version V2 avec Occ_nvs+Liq_vs est très semblable à ce qui est produit naturellement dans un contexte Occ_vs+Liq_vs. Ces résultats suggèrent que, dans ce type de séquence, c'est le trait de la consonne assimilée qui porte l'information phonétique et non la consonne censée porter ce trait. Autrement dit, dans ce cas, l'information portée par un phénomène de variation est prépondérante par rapport à l'information canonique.

Dans une deuxième expérience, la version originale des stimuli a été conservée (V1) tandis qu'une version modifiée (V2') était créée en extrayant l'occlusive des stimuli de V1 (table 3.g). On avait donc quatre types de stimuli:

Voiced stop		Unvoiced stop	
V1	V2'	V1	V2'
glace	(g)las	classe	(c)lasse
g _l as	l _l as	k _l as	l _l as

Table 3.g: Version originale et manipulée des stimuli en fonction du trait de voisement de l'occlusive et de la réalisation allophonique de la liquide

Dans une tâche de choix forcé portant sur les mots, les sujets devaient déterminer s'ils entendaient la version voisée ou non voisée de la paire (*glace* ou *classe*). Les résultats confirment nettement la première expérience: l'occlusive n'est pas nécessaire pour l'identification du trait de voisement de la paire (figure 3.g). Les réponses et temps de réaction pour la version /l_las/ sont même identiques à la version V1. Pour la version /k_las/, on note des résultats moins nets, comme dans la première expérience: on observe surtout l'hésitation des

sujets au travers des TR. Toutefois, même dans cette condition, les sujets identifient clairement le mot grâce au trait de voisement de la liquide.

Les tâches proposées concernaient l'identification globale de l'unité lexicale, ce qui peut favoriser des réponses catégorielles dans le sens où un effet top-down peut influencer le choix des sujets. Nous avons donc proposé une tâche de détection de phonèmes: les sujets devaient détecter l'occlusive dans la même série de stimuli (V1 et V2'), sachant que dans la version V2', l'occlusive était absente.

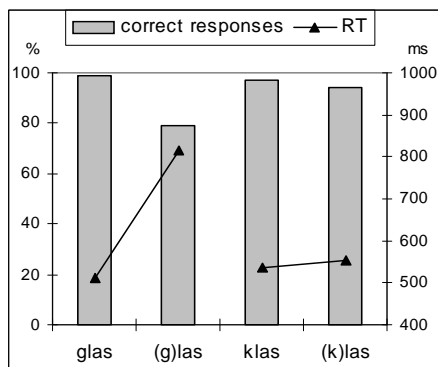


Figure 3.g: Correct responses (left axis in percentage, gray bars) and reaction times (right axis in milliseconds, black points) in the forced choice experiment.

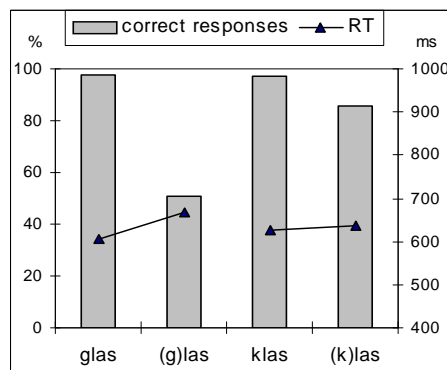


Figure 3.h: Correct responses (left axis in percentage, gray bars) and reaction times (right axis in milliseconds, black points) in the phoneme detection experiment.

Des résultats très similaires ont été observés (figure 3.h): il semble que l'absence de l'occlusive ne soit pas un obstacle à son identification, qui est faite essentiellement grâce aux indices présents dans la liquide.

Nous sommes donc en mesure de conclure que des variations allophoniques non distinctives sont en mesure d'altérer l'identification des mots. Il semble même que la modification du trait de voisement dans les liquides soit l'indice majeur permettant d'identifier le trait de voisement du groupe consonantique et donc du mot. Le fait que le trait de voisement ne soit pas distinctif pour les liquides, donne probablement à cette variation un poids perceptif supérieur à celui qu'elle prend pour les phonèmes (occlusives ou fricatives) pour lesquels le voisement est distinctif. Par conséquent, des modifications non distinctives des traits phonémiques jouent un rôle majeur dans la perception de mots.

Contribution: **Coarticulation ou assimilation?**

Niebuhr O, **Meunier** C. (2011) "The Phonetic Manifestation of French /s#j/ and /j#s/ Sequences in Different Vowel Contexts: On the Occurrence and the Domain of Sibilant Assimilation.", *Phonetica*, Volume 68, No 3, 133-160.

Niebuhr, O., Clayards, M., **Meunier**, C., Lancia, L. (2011) "On place assimilation in sibilant sequences—Comparing French and English", *Journal of Phonetics*, Volume 39, Issue 3, July 2011, Pages 429-451.

Niebuhr, O., Lancia, L., **Meunier**, C. (2008) "On place assimilation in French sibilant sequences". In *Proceedings of the 8th International Seminar on Speech Production (ISSP08)*. 8-12 December 2008, Strasbourg, France, 221-224.

Dans le cadre du projet européen Sound to Sense, j'ai été amenée à travailler avec O. Niebuhr sur le phénomène d'assimilation de lieu (ou de coarticulation) dans des groupes d'obstruantes en français et en anglais. Ce travail visait à déterminer l'utilisation perceptive de détails phonétiques fins (FPD, *Fine Phonetic Details*, Hawkins, 2003) chez des sujets de langues différentes pour lesquelles ces détails phonétiques n'ont pas la même pertinence. L'assimilation de voisement est un phénomène pertinent en production comme en perception en français. En revanche, si le cas de l'assimilation de lieu d'articulation est attestée en anglais (*bad girl* sera produit /baggʌrl/) ça n'est pas le cas pour le français. Ce cas était donc étudié ici dans le cadre d'expérimentations comparatives entre le français et l'anglais.

Le phénomène d'assimilation est souvent considéré comme un changement total des traits d'un phonème. Or on constate que très souvent, même si le cas est attesté dans une langue, les indices observables sur le signal de parole sont, eux, souvent très graduels.

Deux corpus ont été enregistrés en parallèle, l'un pour le français, l'autre pour l'anglais. Dans les deux langues des séquences de fricatives ont été enregistrées dans un contexte CVC#CVC. Les séquences C#C pouvaient être /j#s/ ou /s#j/ et étaient intégrées au sein de phrases porteuses du type: *C'est une trousse chargée* ou *Elle remâche sa viande* (pour le français) et *It's great how the glass shines* ou *The dog needs a brush sometimes* (pour l'anglais). Les trois voyelles /a i u/ étaient manipulées dans la position pré- ou post- séquence. Quatre locuteurs natifs de chaque langue ont enregistré 24 phrases de ce type comprenant les séquences /sʃ/, /ʃs/, /zʃ/ et /ʃz/ dans lesquelles le voisement était donc également manipulé. Les analyses ont porté sur la durée et l'analyse du centre de gravité (CoG) spectral de la totalité de la séquence.

Concernant la durée, les locuteurs français ne montrent pas de hiérarchie entre les deux consonnes, chacune occupant globalement la moitié de la totalité de la séquence. Les locuteurs anglais, en revanche produisent la première consonne toujours plus courte que la seconde, quelque soit son lieu d'articulation et son mode de voisement.

Concernant les analyses spectrales, on constate que l'assimilation de lieu est présente dans les deux langues. Elle se manifeste d'ailleurs graduellement dans ces deux langues, la gradualité

étant toutefois plus marquée pour le français. La gradualité ne semble donc pas être un indice prédominant pour attester le phénomène d'assimilation dans une langue. La différence entre les deux langues se manifeste donc ailleurs. Les locuteurs anglais produisent une assimilation régulière (même si elle n'est pas systématique) qui est strictement régressive (vers les postalvéolaire). Les quatre locuteurs anglais montrent le même pattern de production avec certaines nuances (S4-EN montre une influence postalvéolaire plus forte, figure 3.i).

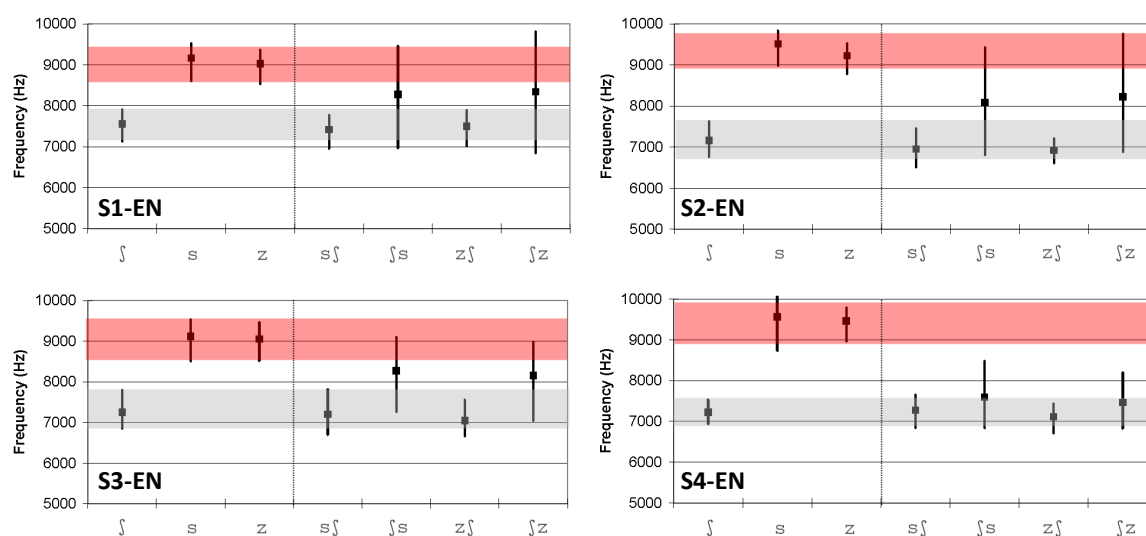


Figure 3.i: Moyenne des CoG pour les fricatives de référence (à gauche de chaque schéma) et les séquences de fricatives pour les 4 locuteurs anglais. Les barres horizontales indiquent approximativement l'aire du CoG pour la post-alvéolaire (en gris) et pour l'alvéolaire (en rouge) de référence.

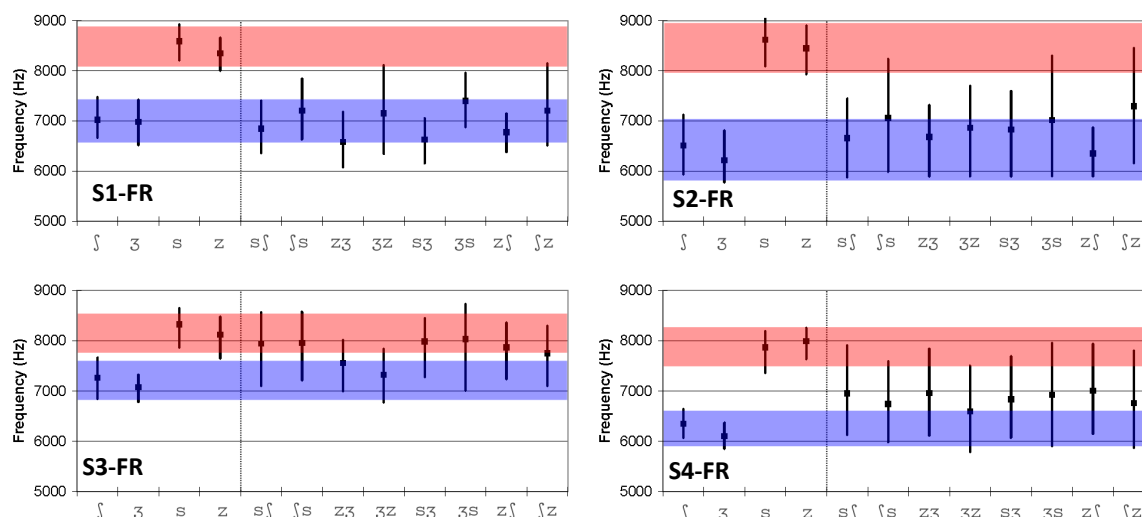


Figure 3.j: Moyenne des CoG pour les fricatives de référence (à gauche de chaque schéma) et les séquences de fricatives pour les 4 locuteurs français. Les barres horizontales indiquent approximativement l'aire du CoG pour la post-alvéolaire (en bleu) et pour l'alvéolaire (en rouge) de référence.

Pour le français, les analyses spectrales montrent une assimilation de lieu de /s/ vers /ʃ/ qui peut être progressive ou régressive. L'assimilation est donc présente mais elle se manifeste moins

régulièrement (elle est notamment plus dépendante des locuteurs). En outre, elle n'est pas conditionnée par la direction de l'assimilation comme en anglais (assimilation régressive), mais plutôt par le lieu d'articulation d'une des deux consonnes (la postalvéolaire, figure 3.j).

Ainsi, pour le français l'assimilation serait plutôt gouvernée par le trait (lieu d'articulation: postalvéolaire) et minoritairement par la direction. A l'inverse, l'assimilation en anglais est gouvernée par la direction et minoritairement par le trait (postalvéolaire).

En conséquence, l'assimilation de lieu est bien présente dans les deux langues sur le plan phonétique. Dès lors, pourquoi l'assimilation de lieu est-elle attestée en anglais comme un phénomène phonologique, ce qui n'est pas le cas pour le français? Une des raisons pourrait être la régularité du processus phonétique. Le phénomène d'assimilation n'est pas systématique pour le français alors qu'il est très régulier pour l'anglais. La régularité du phénomène (présence, sens, etc.) pourrait donc avoir une influence sur l'intégration perceptive de ce trait chez les locuteurs d'une langue. De même, le sens de l'assimilation pourrait également jouer un rôle important dans la représentation des séquences assimilables. L'assimilation de voisement est attestée en français et le sens de cette assimilation est très majoritairement gouverné par la direction (régressive) et secondairement par le trait de voisement (dévoisements plus fréquents). On peut donc faire l'hypothèse que, pour être attestée dans une langue (c'est-à-dire intégrer dans le processus de production et de perception des locuteurs), le phénomène d'assimilation doit être marqué par la régressivité, autrement dit, l'anticipation des gestes articulatoires.

III.2.2 Phénomènes de réductions dans la parole

La parole spontanée fait apparaître des séquences phonétiques très spécifiques encore peu ou pas abordées. Le phénomène « phare » de l'étude des corpus de parole non préparée est *la réduction*¹. "*La réduction est un affaiblissement phonétique des phonèmes [...] Elle renvoie à une diminution de l'amplitude et/ou de la durée des déplacements gestuels, et plus probablement de l'effort articulatoire*" (Meynadier, 2013). Sous ce terme on entend donc aussi bien les élisions que les altérations des segments phonétiques. Johnson (2004) rapporte que dans le corpus de conversation en anglais américain (*The buckeye corpus*, Pitt et al. 2005) l'élision d'un phonème est présente dans 20% des mots et celle de deux phonèmes dans 5% des mots. Plus de 60% des

¹ On notera, sous cette thématique, la tenue d'un Workshop international en 2008 à Nijmegen (Pays-Bas) organisé par M. Ernestus et N. Warner, *The First Nijmegen Speech Reduction Workshop* (<http://www.u.arizona.edu/~nwarner/Workshop.html>), ainsi qu'un numéro spécial de *Journal of Phonetics* (2011, vol. 39).

mots sont altérés sur au moins un phonème et 28% sur au moins deux phonèmes. Toujours dans cette même étude, les catégories lexicales sont différemment affectées puisque 4,5% des mots fonction, contre 6% pour les mots de contenu, perdent au moins une syllabe. Les diverses formes que peut prendre la réduction de la parole sont donc désormais largement décrites dans de nombreuses langues et semblent dépendre de facteurs très variés (Ernestus and Warner, 2011).

Si la réduction semble être présente dans toutes les langues, des travaux montrent que certaines langues pourraient être plus affectées que d'autres. Les mots du danois sont ainsi particulièrement réduits au point que cette langue serait moins intelligible que d'autres langues proches (Gooskens et al., 2010). Plus précisément, certaines contraintes phonotactiques propres aux langues peuvent entraîner un taux de réduction plus important. Notamment, les syllabes ouvertes semblent plus propices à la réduction que les syllabes fermées, ce qui favoriserait la réduction syllabique dans les langues où ce type de syllabe est fréquent (Burchfield and Bradlow, 2014). Enfin, dans une même langue, tous les locuteurs ne réduisent pas avec la même amplitude dans un même style de parole. La réduction est donc présente dans toutes les langues et chez tous les locuteurs mais sa présence reste dépendante de divers facteurs.

Alors que le terme *variation* évoque un phénomène global pouvant affecter tous les types de parole, la *réduction* affecte plus spécifiquement la parole spontanée. Le terme de *réduction* implique une diminution de l'information phonétique, compensée par l'information contextuelle très riche présente en parole spontanée. La réduction phonétique peut alors prendre des caractéristiques très éloignées de ce que l'on trouve en parole lue. Cette forme de variation est différente de celles mentionnées plus haut. Il ne s'agit plus de variations dues au contexte phonétique direct, mais à l'ensemble des propriétés linguistiques et communicationnelles associées à une parole en interaction.

Concernant la réduction, il nous faut répondre à deux questions. **Comment** (quelles formes) et **Pourquoi** (quelles causes). Nous aborderons ici la forme et la nature des réductions de la parole. Les causes possibles seront abordées dans la partie suivante "IV. S'adapter". Notons toutefois que les réductions dont nous parlerons dans cette partie sont, pour l'essentiel, caractéristiques de la parole spontanée et ne sont produites que très rarement en parole lue.

Contribution: **Les formes de réduction**

Meunier, C. (2012) " Contexte et nature des réalisations phonétiques en parole conversationnelle ", Actes des Journées d'Etude sur la Parole, Grenoble (France), juin 2012, 1-8.

La production des sons en parole conversationnelle laisse apparaître des phénomènes de réduction dont les implications phonétiques, articulatoires et perceptives sont encore à éclairer. Face à la diversité des phénomènes de réduction, j'ai tenté d'en distinguer quelques types¹:

Elisions et réductions

Une analyse de ces phénomènes dans le corpus du CID nous a permis de mettre en évidence plusieurs types de caractéristiques. Les phénomènes de réduction se définissent par l'absence de certaines caractéristiques phonétiques par rapport à ce que l'on définit comme forme canonique, prototypique du phonème. Or, cette absence peut revêtir plusieurs formes:

1. **L'élision** d'un ou de plusieurs phonèmes
2. le changement des indices phonétiques dû à une cible articulatoire non atteinte (**undershoot gestures**)
3. la **fusion** (ou coalescence, métaplasme) des segments phonétiques

Le premier cas est le plus connu car il regroupe des phénomènes étudiés depuis très longtemps comme la chute de /ə/ en français. Ces élisions, très phonologisées sont bien connues et leur fonctionnement largement décrit (voir Côté & Morrison, 2007). En revanche, on trouve fréquemment dans les corpus de parole spontanée des élisions moins documentées et plus difficiles à identifier comme le /e/ de *c'était* produit /stɛ/ (Torreira and Ernestus, 2011), *tu vois* produit /tywa/ ou encore *tu sais* produit /tse/ (Meunier, 2012). Quoi qu'il en soit, les élisions (i.e. le phonème omis est clairement identifié) sont souvent identifiables car elles relèvent soit du système phonologique de la langue (schwa), soit de stéréotypes ancrés dans la forme phonologique du lexique (exemplaires).

Les deux autres formes sont plus complexes à appréhender car moins documentées et moins identifiables. Par ailleurs, leur fonctionnement est graduel et leur niveau de dépendance plus difficile à déterminer. Dans le cas des cibles non atteintes, on entend que des phonèmes ne sont pas réalisés dans leur forme canonique. Les phénomènes les plus connus sont la centralisation des voyelles, la non réalisation de l'explosion pour les occlusives ou encore les réalisations approximantes des constrictives.

¹ Cette distinction a pour but de discuter de processus phonétiques complexes. Toutefois ces trois types de réduction pourront être trouvés conjointement dans de nombreuses séquences.

Le phénomène de coalescence (ou métaplasmes) est le moins bien documenté car probablement le plus difficile à appréhender, raison pour laquelle il me semble important d'y accorder de l'attention. Nous entendons par coalescence une *fusion* entre plusieurs segments qui rend impossible l'identification des segments préservés ou omis (figure 3.i).

Notre représentation binaire et paradigmatique des séquences de phonèmes rend l'analyse de ces phénomènes très complexe. Notre mode de représentation lui-même fait défaut: comment annoter une suite de phonèmes lorsqu'on ne sait pas ce qui est produit. Seule notre connaissance de ce qui a été produit (par effet top-down) nous permet une décomposition en phonèmes. Certains de ces phénomènes concernent des formes dites *figées* et sont assez stéréotypés. Ils sont alors aisément identifiables car reproductibles. C'est le cas de la forme canonique 'je ne sais pas' /ʒənəsɛpa/, très fréquemment réalisée en conversation [ʃɛpa]. De même certaines formes peuvent être stéréotypées chez un locuteur. La séquence *tu vois* /tyvwɑ/ est très fréquemment réalisée [tʃɑ] chez un des locuteurs du CID, par exemple. Mais ces formes sont, d'une certaine façon, attendues car elles affectent des séquences de mots apparaissant souvent dans la conversation selon le même patron (formes *figées*).

Toutefois, la réduction affecte également des séquences de mots qui ne sont pas toujours reproductibles dans la conversation. On trouve ainsi des séquences non identifiables sous forme de suite d'unités phonétiques sur le signal de parole alors qu'elles sont perceptibles dans un contexte élargi. La caractéristique de ces séquences, à la différence des phénomènes de réduction typique comme les élisions claires ou les altérations (du type assimilations), est l'impossibilité d'identifier les phonèmes réalisés et ceux qui sont omis (figures 3.l & 3.n). La séquence apparaît comme une sorte de fusion syntagmatique des éléments qui la composent. Cette fusion peut se produire en interne de mots comme aux frontières¹.

Une première hypothèse consiste à considérer qu'il s'agit juste d'une réduction de l'effort articulatoire répondant à des contraintes physiologiques dont la conséquence est une très forte sous-spécification du niveau phonétique lorsque la redondance de l'information linguistique le permet. Dans cette hypothèse, ces productions ne contribuent pas à l'information. La compréhension du message serait alors garantie par des informations descendantes permettant de pallier la sous-spécification phonétique (Warren & Obusek, 1971).

¹ Notamment, une grande partie des productions lexicales dans ce corpus étant des mots fonction monosyllabiques, ces fusions apparaissent souvent dans des suites de monosyllabes.

L'observation des réductions présentes dans le CID nous amènent à soutenir une autre hypothèse: certains de ces phénomènes seraient régis par des contraintes physiologiques mais répondraient également à des contraintes du système linguistique. Dans plusieurs cas, nous avons pu noter que le processus de réduction tendait à préserver certaines caractéristiques phonétiques porteuses d'information.

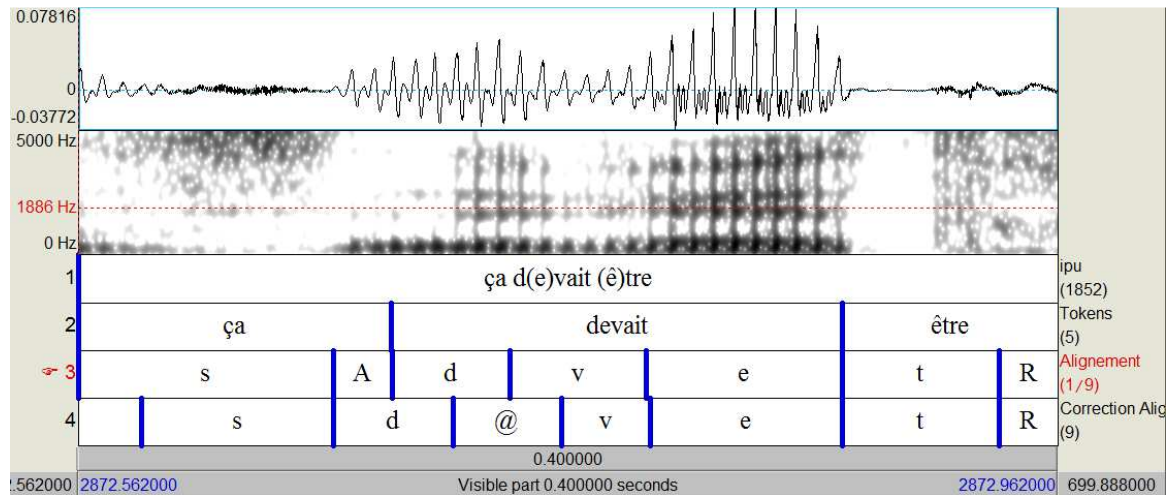


Figure 3.k: Séquence transrite "ça d(e)vait (ê)tre" (niveau IPU). La segmentation est proposée par un expert humain. Les niveaux "Tokens" et "Alignement" sont produits automatiquement à partir de la transcription. Le 4^{ème} niveau propose une correction de l'alignement en fonction de l'interprétation d'un expert humain.

Par exemple, dans la figure 3.k, le /a/ transcrit n'a, en fait, pas été réalisé par le locuteur. La production effective est /sdvæt/ (voir correction alignement). Habituellement, dans ce contexte, on devrait trouver une assimilation de voisement entre s et d, ce qui n'est pas le cas. La perception correcte de la séquence serait alors préservée par cette absence d'assimilation qui agit comme une empreinte de la présence sous-jacente du /a/. De même, sur la figure 3.l, on constate une réalisation assez inhabituelle de la suite d'occlusives /bd/ due à l'élision de /u/ dans *au bout d'un*. En effet, dans un contexte de réduction très forte, on aurait pu s'attendre à ce que l'élision de /u/ entraîne une fusion des deux occlusives ou, au moins, une réduction de leurs composantes. Or, on constate que l'explosion de la bilabiale (pourtant habituellement assez faible, voire inexistante) est clairement réalisée. Notre hypothèse est que, dans de nombreux cas, ces phénomènes de coalescence, contrairement aux réductions *phonologiques*, ne correspondent pas systématiquement à l'élision simple d'un segment mais sont caractérisées par des processus articulatoires complexes et variés tendant à préserver certains indices pertinents qui rendent accessible l'information phonétique, et donc le message linguistique. Ces processus articulatoires représenteraient une sorte de **socle phonétique**. Ce socle fournirait des informations phonétiques minimales permettant, en partie, à l'auditeur d'interpréter le message

linguistique. Notons que dans les deux exemples mentionnés, on se trouve face à des configurations assez étonnantes et peu répertoriées en parole contrôlée: un enchaînement de segments phonétiques prototypiques sans assimilation ou coarticulation, là où l'on s'attendrait à en trouver. Le blocage des processus d'assimilation ou de coarticulation pourrait préserver une information phonétique sous-jacente.

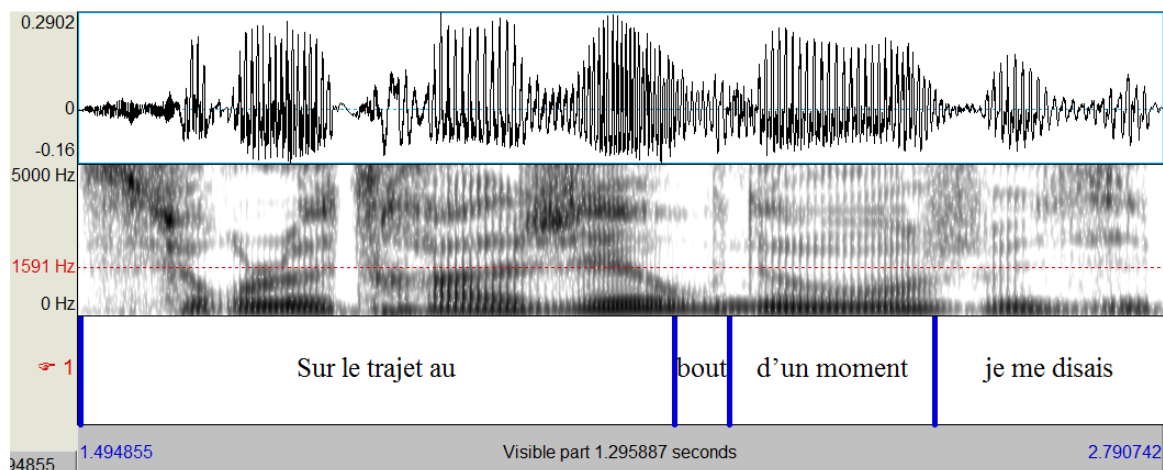


Figure 3.1: Corpus CID, locuteur YM. Forte réduction de la séquence *un moment* et *je me disais*. Dans la séquence *bout d'un*, l'élision du /u/ n'entraîne pas d'assimilation entre les deux occlusives.

Les aligneurs, comme les phonéticiens, ne proposent pas de solutions satisfaisantes pour l'identification de ces séquences dans la mesure où les deux approches reposent sur une représentation phonétique discrète et paradigmatique du signal de parole. Or, les fusions ne peuvent être décrites en ces termes, ce qui soulève des questions théoriques très intéressantes. Un des objectifs des phonéticiens pourrait être de fournir une interprétation articulatoire mais aussi linguistique de ces phénomènes de réduction tout en proposant des solutions pour une annotation phonétique adéquate.

La tendance est forte de voir dans ces formes de réduction des contraintes purement physiologiques dues à l'augmentation du débit de la parole. Sans exclure cette explication, mon hypothèse est que ces phénomènes sont probablement à relier avec les propriétés de secteurs linguistiques de plus haut niveau comme le discours, l'interaction, la prosodie, etc. Ils n'apparaissent pas n'importe où dans la conversation (Adda-Decker, 2008) et des détails phonétiques ténus (Hawkins, 2003) permettent la préservation et l'intelligibilité de la chaîne parlée.

Contribution: **Elisions et fusions**

Meunier, C. (2013) "Phoneme deletion and fusion in conversational speech", ExAPP 2013 Experimental Approaches to Perception and Production of Language Variation, Copenhagen (Denmark), mars 2013.

Une présentation à Copenhague en 2013 m'a donné la possibilité d'aller un peu plus loin dans la comparaison des formes de réduction en essayant d'en quantifier la fréquence d'apparition et d'en présenter une description plus précise. J'ai donc mené une étude portant sur une comparaison de deux types de réduction présentés plus haut, l'*élision* et la *fusion*. Elles peuvent être caractérisées ainsi:

Elision (Figure 3.m):

- Au moins un phonème n'est pas réalisé
- Nature catégorielle du phénomène
- L'absence du(es) phonème(s) est clairement identifiée

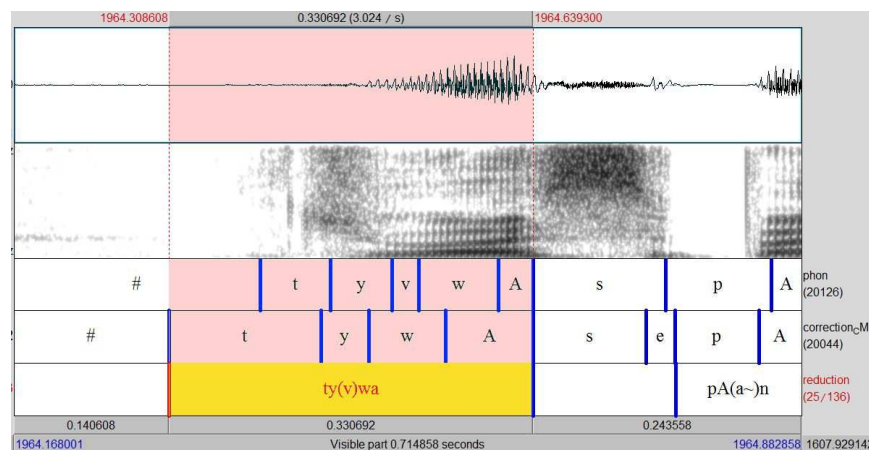


Figure 3.m: Exemple d'élision. Séquence *Tu vois je sais pas*. Dans la séquence *tu vois*, le /v/ n'est pas réalisé /tywa/.

Fusion (Figure 3.n):

- Au moins deux phonèmes sont fusionnés
- Nature graduelle du phénomène
- Les phonèmes fusionnés ne sont pas individualisables

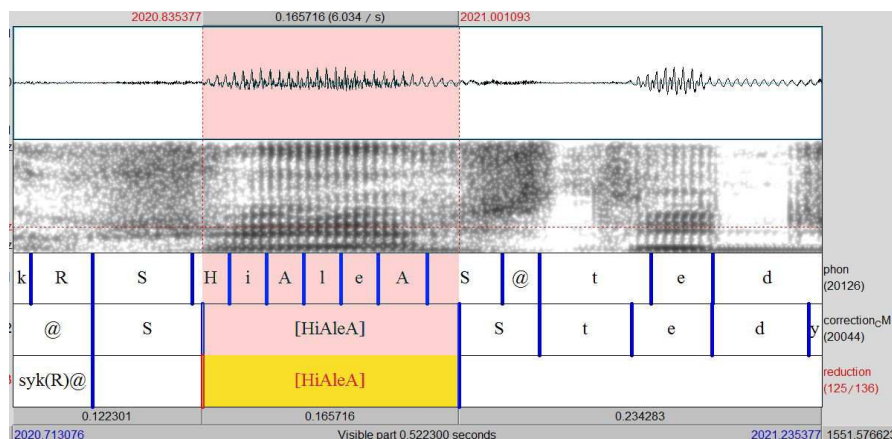


Figure 3.n: Exemple de fusion. Séquence *Je suis allé acheter*. Dans la séquence /ʃialea/ les phonèmes ne sont pas identifiables ni distinguables.

Deux locuteurs dans le corpus du CID (Meunier, 2013) ont été analysés sur 200 secondes chacun. Les 1322 phonèmes produits sur ces 200 secondes ont une durée moyenne de 72ms. Le nombre de mots produit est de 535. Le nombre de cas d'élisions est de 52 alors que 69 cas de fusions ont été relevés. Nos résultats ont pu mettre en évidence que les deux phénomènes réunis réduction (fusion + élision) affectent considérablement les phonèmes (18.5%) et les mots (30%) de notre échantillon. Toutefois, le phénomène de fusion est plus fréquent que l'élision (figure 3.o).

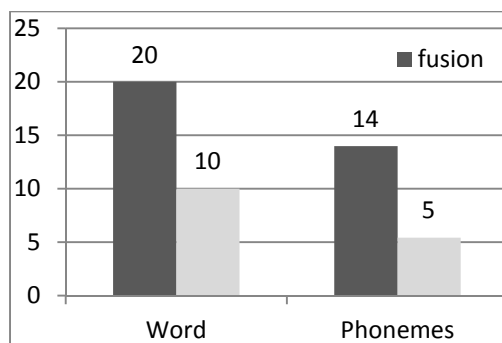


Figure 3.o: Pourcentage de mots et de phonèmes affectés, respectivement, par la fusion ou l'élision.

Bien entendu, chaque phénomène de fusion affecte plus de phonèmes que l'élision dans la mesure où la fusion suppose le contact d'au moins deux phonèmes alors que l'élision peut ne concerner qu'un seul phonème. Nos observations montrent que l'élision affecte majoritairement un seul phonème (83% des cas) tandis que la fusion concerne majoritairement deux phonèmes (2 phonèmes: 59% des cas; 3 phonèmes: 24%; 4 phonèmes et plus: 17%). La fusion semble donc se répartir sur des séquences bien plus larges que l'élision.

Les voyelles et consonnes sont affectées de façon comparable chez les deux locuteurs aussi bien pour l'élision que pour la fusion. Elision et fusion affectent majoritairement les mots monosyllabiques (66%), et bisyllabiques (28%) dans une moindre mesure. De même, ces deux

formes de réduction se retrouvent plus souvent dans les mots de fonction (71%). Toutefois, cette répartition de la réduction dans les mots de fonction mono- ou bi-syllabiques ne signifie pas que ces mots sont particulièrement affectés, ils sont juste les plus fréquents dans les corpus. Notons que les mots de fonction sont plus affectés par (81%) que par la fusion (61%). Ces résultats doivent, malgré tout, être confirmés par un nombre plus important de données. Nous avons en effet noté certaines différences entre nos deux locuteurs, dont l'un produit plus de réduction.

Enfin, un dernier résultat nous semble intéressant. Si, dans la quasi-totalité des cas, l'élision est contenue dans un seul mot (98%), la fusion se répartit majoritairement sur plusieurs mots (deux mots au moins: 60%). Ce résultat n'est pas si évident qu'il y paraît. En effet, le phénomène de fusion, même s'il affecte plusieurs phonèmes, pourrait être réparti à l'intérieur d'un seul mot, ce qui n'est pas le cas, la plupart du temps.

Ce dernier constat vient confirmer une hypothèse concernant le niveau de dépendance de ces deux phénomènes de réduction. L'élision affecte le plus souvent un seul phonème à l'intérieur d'un seul mot (cette élision est le plus souvent de type stéréotypée). En revanche, la fusion affecte le plus souvent deux phonèmes ou plus répartis sur au moins deux mots. L'élision semble ainsi être dépendante de la structure phonologique du mot (exemplaires), tandis que la présence de fusion témoignerait d'une dépendance de plus haut niveau, telle que l'organisation prosodique ou discursive.

L'analyse des phénomènes de réduction (notamment *fusion*) est toutefois extrêmement complexe car ils sont non reproductibles (chaque séquence semble unique concernant le contexte phonétique et les unités lexicales impliquées), peu accessibles d'un point de vue perceptif et peu adaptés aux analyses acoustiques (certains gestes articulatoires pourraient avoir un rôle perceptif important sans laisser d'indices acoustiques interprétables). Il semble donc nécessaire d'envisager les analyses phonétiques au travers de méthodologies complémentaires telles que le traitement automatique de la parole, l'expertise phonétique, la prise en compte d'autres niveaux linguistiques mais aussi l'enregistrement de données articulatoires.

III.3 BILAN

Dans cette partie il était question de faire le lien entre les unités des systèmes phonologiques et leur usage. Nous avons vu que les sons ne sont pas utilisés de façon homogène dans l'usage habituel de la langue. Certains sont plus fréquents, d'autres plus rares. En revanche, les gestes articulatoires représentés dans l'inventaire phonologique du français sont relativement bien préservés dans l'usage de la langue, ce qui tendrait à conforter l'idée d'un équilibre dans l'utilisation de nos gestes dans la parole. Cet aspect demande à être conforté par des études similaires portant sur d'autres langues, ce que j'envisage de développer (voir *Perspectives* à la fin de ce document).

Il était également question de donner une description des formes de variation et de réduction phonétiques que l'on peut trouver dans la parole. J'ai distingué les variations classiques (assimilation et coarticulation), que l'on trouve dans tout type de parole, des réductions que l'on trouve plus spécifiquement dans la parole spontanée. De plus, j'ai distingué plusieurs formes de réduction (élisions, fusions, etc.). Cela fait évidemment beaucoup de catégories de variations possibles. Or, si l'on regarde le signal de parole, cette catégorisation n'est pas aussi claire.

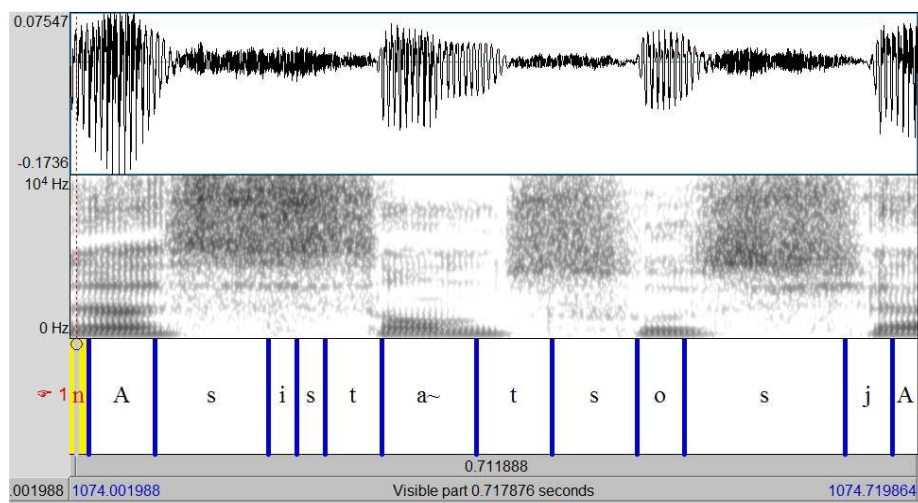


Figure 3.p: Locutrice du CID (ML) produisant les mots *assistante sociale*. La séquence /sist/ présente une forte réduction.

Sur la figure 3.p, on constate que la séquence /sist/ est réalisée sous la forme d'une longue fricative. En quoi consiste cette forme de réduction? Une assimilation (du /t/, du /i/)? Une fusion (de tous les phonèmes)? Une élision (du /t/ et du /i/)? Peut-être les trois en même temps. Mais comment savoir, de toute façon? Ces catégories sont-elles pertinentes? Nous en avons probablement besoin pour nous représenter ces formes très complexes. Toutefois, il me semble que toutes les formes de réduction (élision, assimilation, fusion, phénomènes d'undershoot)

relèvent d'un processus global de diminution de l'effort articulatoire. Ceci étant dit, on peut supposer que cette diminution est gouvernée par différents secteurs linguistiques. Les propriétés phonologiques d'une langue gouverneraient plutôt l'assimilation, tandis que l'élision de phonème semble être plus souvent dépendante de propriétés lexicales. La fusion de plusieurs phonèmes (recouvrant plusieurs mots) serait, en revanche, dépendante de secteur linguistiques couvrant des domaines plus larges. Ce qui pourrait expliquer que toutes ces formes se retrouvent, parfois indistinctement, dans la parole spontanée. Alors qu'en lecture de mots isolés, seuls les phénomènes d'assimilation ou, plus rarement, d'élision sont présents.

L'analyse des formes très réduites (fusions) de la parole m'a par ailleurs montré les limites de l'information qu'apporte l'analyse du signal de parole. Il est difficile, sans information de type articulatoire, de comprendre comment fonctionne le système de production au cours de ces séquences: s'agit-il simplement de réduction extrême où l'information phonétique est totalement sous-spécifiée (donc peu informative)? Ou bien y'a-t-il dans ces phénomènes de réduction une stratégie de maintien d'information minimale (socle) permettant un accès à l'information? Un travail d'analyse articulatoire de ces séquences permettrait de mieux interpréter la nature de ces formes de réduction (voir *Perspectives* à la fin de ce document).

IV. S'ADAPTER

Parler fait référence à de nombreuses notions et, selon la situation et l'interlocuteur, ce terme pourra prendre des significations très variées: quand le phonéticien pensera à la production de la parole, d'autres y verront communication (activité sociale) ou libération (thérapeutique). Mais parler c'est aussi donner de nombreuses informations sur soi, son âge, son sexe, parfois son origine géographique, mais aussi ses caractéristiques socio-culturelles ou encore son état émotionnel. Parler est ainsi une activité plaisante pour certains et pesante pour d'autres. Parler c'est aussi construire la réalité au moment où l'on parle. Indépendamment du message lui-même, on parle souvent pour séduire ou, tout au moins dans l'intention d'avoir un impact sur nos interlocuteurs (rire, compassion, etc.). Parler est donc aussi une stratégie. A l'inverse, l'intention communicative est parfois plus complexe à cerner, lorsque le locuteur soliloque par exemple. Et puis il est des paroles dont le contenu linguistique est sans importance, comme dans les activités dans lesquelles l'émission sonore suffit (mots échangés pour la cohésion du groupe dans des sports collectifs, ou mots produits pour calmer ou endormir un enfant, etc).

Tout cela c'est parler, produire des sons du langage. Il faut bien admettre alors que les objets linguistiques et les contextes sur lesquels nous travaillons, nous linguistes, sont vraiment très réduits. Toutefois, tout cela représente la parole, une même parole, une même fonction fondamentale, avec les mêmes outils. Mais une parole très **flexible**.

Une des capacités fondamentales des êtres vivants est probablement celle de **l'adaptation**. A ce titre, les mécanismes de production et de perception de la parole y sont au cœur. D'une certaine façon, on pourrait dire que parler c'est, en partie, adapter en permanence le contrôle moteur des organes de la parole en fonction des conditions dans lesquelles on se trouve, comme le montre des travaux récents sur l'imitation (Dohen et al., 2010). C'est d'ailleurs probablement une caractéristique du fonctionnement "normal" de la parole (voir plus loin "Aux contraintes").

Lindblom & Lindgren (1985) observent dans un corpus de parole spontanée que la notion d'invariant acoustique ne peut s'appliquer aux productions naturelles des locuteurs: les conditions naturelles de parole impliquent une variabilité phonétique telle qu'il est impossible d'y trouver une quelconque invariance acoustique. La recherche infructueuse d'invariants physiques a conduit Lindblom (1990) à proposer un modèle adaptatif de la production et la perception de la parole (*Hypo and Hyper Speech Theory*). Ce modèle ne postule plus l'existence d'un invariant physique. La communication parlée doit répondre à deux exigences: 1/ le besoin

d'intelligibilité pour l'auditeur et 2/ le principe d'économie articulatoire pour le locuteur. Ainsi, le locuteur produit son énoncé selon un principe de *variabilité adaptative* de façon à ce que les deux exigences (intelligibilité et économie articulatoire) soient respectées. L'articulation est soutenue si la situation l'impose et, selon un principe d'économie, relâchée si cela est possible sans nuire à la compréhension. Une grande partie de la variabilité de la parole peut donc être expliquée par cette adaptation constante aux exigences de la communication. La théorie de Lindblom postule que le système perceptif procède à l'intégration parallèle de différentes sources d'information : le signal de parole lui-même et l'ensemble des informations contextuelles (sémantiques, pragmatiques, etc.). Cette configuration repose sur l'existence d'un processus interactif du traitement de la parole : le message sonore est traité selon un processus ascendant (*bottom-up*) du signal acoustique vers les unités linguistiques mais également descendant (*top-down*) dans la mesure où les informations de plus haut niveau (sémantique, pragmatique), concourent à la compréhension du message.

Mes travaux se situent pleinement dans la théorie H&H de Lindblom. Toutefois, concernant la production phonétique, je fais l'hypothèse d'une adaptation complexe mettant en jeu à la fois les contraintes communicationnelles, mais également des contraintes linguistiques (interaction des différents secteurs linguistiques comme le lexique, la syntaxe, etc.) et des contraintes physiologiques.

Je présenterai dans cette partie différents aspects de cette fonction d'adaptation en lien avec trois axes. Il s'agit des langues, du type de parole et des contraintes physiologiques (dues à la pathologie). Bien sûr, ces trois axes reflètent mes travaux, mais la structure aurait pu être différente. Les contraintes physiologiques et les langues ont été étudiées en fonction des situations de parole ce qui entraîne parfois des chevauchements entre les parties. Toutefois, ces distinctions me semblent pertinentes pour mettre en place les points clefs de mes travaux :

- Le phénomène d'adaptation dans la production de la parole est une caractéristique du locuteur sain
- cette adaptation est universelle dans son principe, mais spécifique aux langues dans sa forme qui dépend de la structure des systèmes linguistiques
- cette adaptation est affectée lorsque le corps est affecté dans son fonctionnement physiologique
- cette adaptation est affectée lorsque le locuteur n'est pas natif
- enfin, cette adaptation est l'une des causes majeures des formes de variation
- à ce titre les formes de variation sont *la trace du fonctionnement de la parole* et contribuent pleinement au système linguistique.

IV.1 AUX LANGUES

Chaque individu naît et se développe dans un environnement linguistique particulier. Cet environnement peut être monolingue mais il est le plus souvent plurilingue. Il peut être stable ou mouvant, c'est-à-dire que les langues présentes dans l'environnement peuvent changer. Nous sommes tous capables (avec des compétences variées) de nous adapter à une nouvelle langue (Baese-Berk et al., 2013). Mais cette adaptation n'est pas toujours simple car elle nécessite l'apprentissage d'un nouveau système linguistique caractérisé par une syntaxe, une prosodie, des mots différents, mais aussi par un système de sons dont les propriétés articulatoires et perceptives viennent perturber celles du système natif (Kuhl, 1995; Best, et al., 1988). La perception, mais aussi l'acquisition d'une langue seconde, est donc fondamentalement différente d'une situation d'acquisition première du langage (Cutler & Weber, 2007) car elle intervient alors que le système du locuteur/auditeur possède déjà toutes les compétences linguistiques spécifiques de sa langue maternelle. L'auditeur se retrouve alors dans une situation défavorable (*adverse condition*) lorsqu'il est confronté à une langue étrangère (Hazan, 2007, Mattys et al., 2012) et force est de constater que nous ne sommes pas tous égaux dans cette forme d'adaptation, les variations inter-individuelles dans la perception d'une langue étrangère pouvant être considérables (Hazan et al., 2010).

Je développerai dans cette partie deux aspects de cette adaptation: le premier tient en une comparaison de plusieurs systèmes vocaliques et de l'influence des spécificités de chaque système sur la production et la perception des voyelles. Le deuxième concerne le lien production/perception dans le cadre de l'apprentissage d'une langue seconde.

IV.1.1 Systèmes phonologiques et variations phonétiques

Contribution:	Densité des systèmes vocaliques
Meunier , C.; Espesser, R.; Frenck-Mestre, C. (2006) "Phonetic variability as a static/dynamic process in speech communication: a cross linguistic study.", 10th Conference on Laboratory Phonology (2006 juin 29-juillet 1 : Paris).	
Meunier C., Frenck-Mestre C., Lelekov-Boissard T., Le Besnerais M. (2004) "La perception des systèmes vocaliques étrangers: une étude inter-langues", Journées d'Etudes sur la parole, Avril 2004, Fès, Maroc, 377-380.	
Meunier C., Frenck-Mestre C., Lelekov-Boissard T., Le Besnerais M. (2003) "Production and perception of foreign vowels: does the density of the system play a role?", Proceedings of the 15th International Congress of Phonetic Sciences, August 2003, Barcelona, Spain, 723-726.	
Meunier , C.; Espesser, R.; Frenck-Mestre, C. (2005) "The role of temporal information in vowel perceptual areas: a cross-linguistic study.", <i>Plasticity in Speech Perception</i> (2005 juin 15-17 : Londres, UK).	

Ces travaux visent à comprendre comment la *densité* d'un système vocalique d'une langue, c'est à dire le nombre faible ou important de voyelles, joue un rôle dans la production et la perception des voyelles. Chaque langue dispose d'un inventaire de sons différent. La structure de cet inventaire n'est pas aléatoire mais suit des tendances universelles (Vallée, 1994). Plusieurs modèles tentent de rendre compte du fonctionnement perceptif en tenant compte de la diversité des langues et de leur structure phonologique (Kuhl, 1995; Best, 1995). Une hypothèse stipule que les langues à système vocalique réduit pourraient produire des catégories plus vastes que des langues à système vocalique rempli (Manuel & Krakow, 1984). Toutefois, plusieurs études ont produit des résultats contredisant cette hypothèse (Maddieson & Wright, 1995). Notre objectif était donc de vérifier cette hypothèse dans une série d'analyses et d'expériences mêlant production et perception.

Dans une première série de travaux, nous avons pu infirmer une hypothèse stipulant que les espaces de réalisation des voyelles, donc leur variabilité, seraient conditionnés par le nombre de voyelles dans un système. En effet, dans nos travaux, des locuteurs anglais (10 voyelles dans le système) produisent des espaces de réalisations bien plus larges que des locuteurs espagnols (5 voyelles) (figure 4.a).

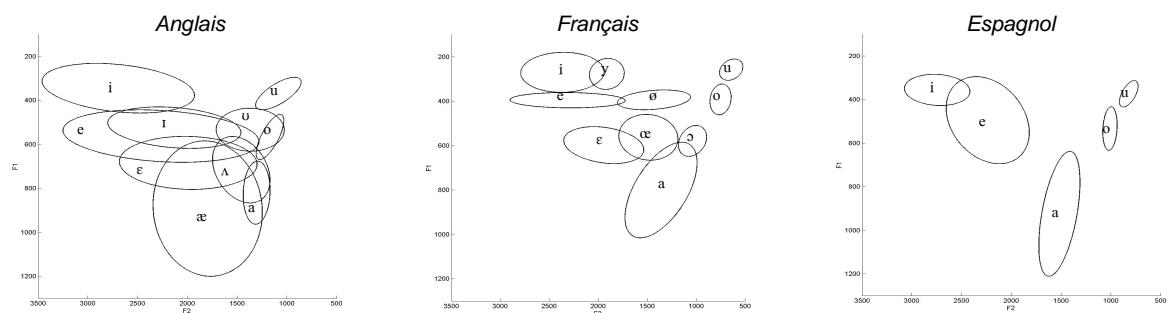


Figure 4.a: espaces de dispersion des voyelles pour les locuteurs des trois langues

Les corpus analysés étaient constitués de prononciation de voyelles isolées. Ces résultats surprenants tendent à montrer que la dispersion de réalisation des voyelles dans une langue donnée serait dépendante de facteurs multiples et complexes faisant intervenir d'autres paramètres que la densité du système.

Au cours d'une expérience comportementale d'identification de voyelles, nous avons demandé à des locuteurs français, anglais et espagnols de transcrire des voyelles présentées auditivement (en premier, les voyelles de leur propre langue, puis celles des deux langues étrangères). Pour chaque langue, chacune des voyelles avait été prononcée 5 fois par 2 locuteurs différents (femme, homme). Ainsi, chaque sujet entendait 10 exemplaires de chaque voyelle.

Nous avons pu constater que les auditeurs anglais montrent de grandes difficultés dans une tâche simple d'identification de voyelles (même dans leur propre langue), à l'inverse des auditeurs espagnols et français pour lesquels la tâche s'est révélée relativement simple (figure 4.b & 4.c; table 4.a). Notre interprétation de ces résultats va dans le sens d'un manque d'information pour les sujets anglais (voir plus loin).

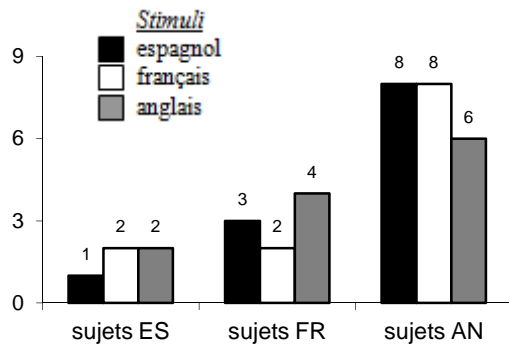


Figure 4.b : Nombre moyen de voyelles identifiées pour chaque stimulus présenté (un nombre important signifie que le sujet attribue plusieurs représentations pour une même voyelle).

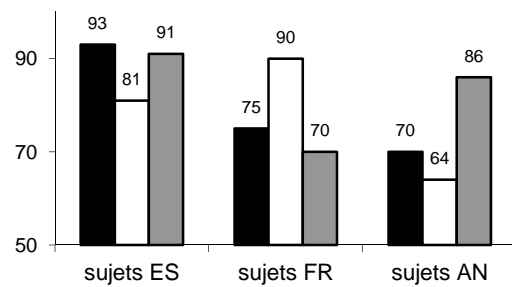


Figure 4.c : Pourcentage moyen d'identification de la voyelle ayant le meilleur score pour chaque langue.

		Voyelles perçues					
		i	e	a	o	u	
Voyelles produites	Voyelles espagnoles	i	78	22			
		e		100			
		a			100		
		o				100	
		u				17	83
	Voyelles françaises	i	97				3
		e	10	90			
		a		1	99		
		o				62	38
		u					100
	Voyelles anglaises	i	97	3			
		e	1	97	2		
		a			100		
		o				99	1
		u				4	96
		Voyelles perçues					
		i	e	a	o	u	
Voyelles produites	Voyelles françaises	i	100				
		e		91			
		a			98		
		o				97	
		u					100
	Voyelles espagnoles	i	22	73			
		e		28			
		a			99		
		o			11	13	
		u				33	61
	Voyelles anglaises	i	89	11			
		e		55			
		a			100		
		o				82	
		u				5	95
		Voyelles perçues					
		i	e	a	o	u	
Voyelles produites	Voyelles anglaises	i	97	1			
		e		96	0,5		
		a	3		80		0,5
		o		1	5	89	2
		u					93
	Voyelles espagnoles	i	66	2	0,5	1	
		e	2	28	0,5		1
		a		0,5	74	0,5	0,5
		o	5		6	77	2
		u	4			7	68
	Voyelles françaises	i	95	0,5		0,5	
		e	6	34	1		
		a	0,5	1	80	0,5	
		o	0,5		2	61	22
		u	0,5	0,5	2	0,5	82

Table 4.a : Pourcentage d'identification des voyelles présentes dans les trois langues /i e a o u/.

Nous avons pu observer une bonne cohérence entre les résultats acoustiques et perceptifs : 1/ concernant la position globale des systèmes, le décalage du système français (plus fermé et plus postérieur) se retrouve dans les résultats des sujets espagnols et anglais; 2/ concernant les espaces de réalisation: les sujets français et espagnols produisent des catégories de voyelles bien distinctes et n'ont pas de difficulté à catégoriser les voyelles. En revanche, la production des anglais montre une forte variabilité et les résultats des expériences comportementales montrent qu'ils éprouvent des difficultés dans la tâche de catégorisation.

Il ressort de ces résultats que les sujets espagnols arrivent, sans hésitation, à catégoriser des stimuli qui n'appartiennent pas à leur système vocalique. On peut donc interpréter ces résultats en observant que, moins il a de voyelles dans un système, plus la catégorisation est aisée.

Toutefois, ces résultats doivent être confirmés par de nouvelles expériences comportementales. On ne peut pas déduire que la facilité de catégorisation des espagnols est due à de larges espaces de perception plutôt qu'à un choix restreint de catégories. Par ailleurs, la difficulté de catégorisation des sujets anglais ne peut pas être due seulement au nombre important de voyelles car les sujets français, pour un même nombre de voyelles, catégorisent plus aisément. Certains travaux ont pu montrer que, au delà de neuf unités, un système vocalique dispose, en général d'un système secondaire d'indices (Vallée, 1994; Schwartz et al., 1994). C'est le cas de l'anglais : les auditeurs anglais utilisent la durée comme indice secondaire leur permettant de distinguer les voyelles. Or, cet indice n'est pas manipulé dans notre expérience. On peut donc supposer que les sujets anglais n'avaient pas assez d'indices pour réaliser correctement la tâche.

En conclusion, l'anglais dispose, d'une part, d'un système vocalique à forte densité et, d'autre part, d'un système d'indices complexe(s) pour la caractérisation des voyelles (durée, accentuation). A l'inverse, la densité du système espagnol est plus faible et ses caractéristiques reposent sur le timbre (F1/F2). Le système français est intermédiaire: il est caractérisé par une forte densité et un système d'indices plus complexe (nasalité) que celui de l'espagnol mais ne dépendant pas de contraintes supra-segmentales comme pour l'anglais. En conclusion, la taille des catégories vocaliques ne semble pas être affectée uniquement par la densité. Ainsi, si la densité des systèmes phonologiques pourrait bien avoir un impact sur la taille des catégories, il semble que cette relation soit plus complexe qu'il ne semblait a priori et nécessite de prendre en compte l'ensemble des caractéristiques liées à la production et à la perception des voyelles.

En conséquence, dans une deuxième série de travaux, nous avons cherché à montrer jusqu'à quel point la durée est utilisée par les anglais dans une tâche de catégorisation sur des continuum synthétiques de voyelles (a-i ; a-u ; i-u). Les trois continuum étaient présentés avec deux durées (200 ms et 400 ms). Il était demandé à des sujets français et anglais de décider si la cible qui leur était présentée sur un écran ("a as in farm") correspondait à la voyelle qu'ils entendaient. Les résultats montrent que les sujets anglais utilisent les indices temporels pour l'identification de la voyelle /a/. Ainsi, lorsqu'il leur est demandé si un stimulus entendu correspond à la voyelle /a/, les sujets anglais répondent différemment selon que le stimulus a une durée courte (200 ms) ou longue (400 ms). En revanche, les sujets français présentent les mêmes patterns de réponses quelle que soit la durée des stimuli (figure 4.d). Ces résultats tendent à confirmer notre hypothèse selon laquelle la grande dispersion des voyelles anglaises en production ainsi que la difficulté que montrent les sujets anglais dans des tâches d'identification de voyelles seraient dues à la spécificité du système vocalique de l'anglais : les

indices temporels, entre autres, semblent nécessaires pour une identification correcte des voyelles.

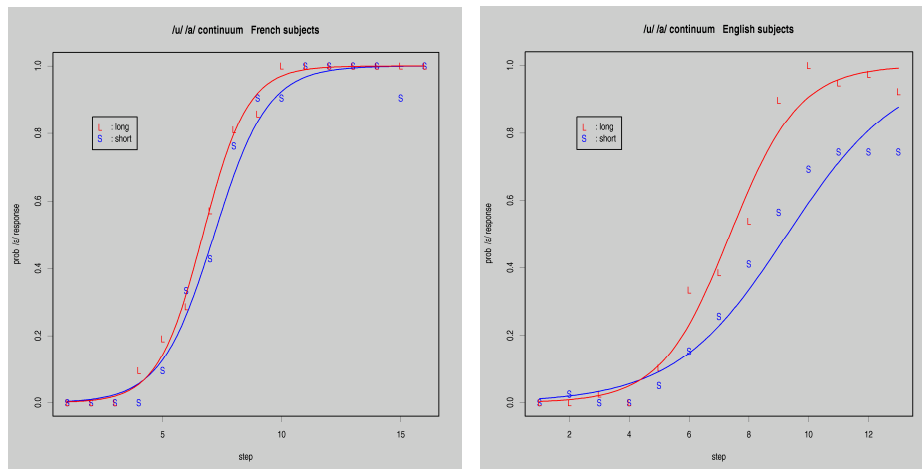


Figure 4.d: Modélisation des réponses /a/ (rouge pour les stimuli longs, bleu pour les stimuli courts) dans un continuum /u/-/a/. A gauche: sujets français; à droite: sujets anglais.

Toutefois, nous n'avons obtenu cet effet que pour la voyelle /a/, ce qui est surprenant. Nous nous attendions, en effet, à ce que la voyelle /i/ soit plus particulièrement affectée par des variations temporelles. La proximité du /i/ (toujours long en anglais *heat*) et du /ɪ/ (toujours bref *hit*) dans le système vocalique de l'anglais aurait dû faire apparaître un effet plus marqué dans cette région du système.

IV.1.2 Le lien perception-production dans l'apprentissage d'une langue seconde

Contribution:	Perception des phonèmes non natifs
<p>Frenck-Mestre, C.; Meunier, C.; Espesser, R.; Holcomb, P.; Daffner, K. (2005) "Perceiving Nonnative Vowels: The Effect of Context on Perception as evidenced by Event-Related Brain Potentials.", <i>Journal of Speech Language and Hearing Research</i>, vol. 48, 1496-1510.</p> <p>Frenck-Mestre, C.; PERI, P.; Meunier, C.; Espesser, R. (2010). ERP evidence of the acquisition of non-native contrasts in late learners. <i>Proceedings of New Sounds 2010</i> (10 : 2010 avril 30-mai 3 : Poznan, POLAND). 2010, p. 119-124. Cederom.</p>	

Ces travaux se situent dans le cadre de la thèse de Pauline Péri que j'ai co-encadrée sur la partie phonétique de son sujet. Le projet de thèse prenait appui sur les données récemment mentionnées ci-dessus et mettant en évidence les différences selon la langue parlée dans la production des voyelles (Meunier et al., 2004) ainsi que dans la perception de celles-ci (Meunier et al., 2005; Frenck-Mestre et al., 2005) et le rôle de l'expérience de l'auditeur/locuteur lors de l'acquisition des voyelles étrangères (Frenck-Mestre et al., 2007; Peri & Frenck-Mestre, 2007).

Si de manière générale les études sur la question suggèrent l'existence d'un lien entre ces deux versants du langage (production et perception), la nature exacte de celui-ci reste à déterminer. La production d'une langue chez un locuteur suppose que, depuis la naissance, se sont mis en place des « routines articulatoires » d'autant plus efficaces qu'elles sont sélectives, c'est à dire que sont privilégiées les commandes articulatoires et musculaires spécifiques à la production de notre langue maternelle, au détriment des sons absents de notre langue. Or, l'apprentissage d'une langue seconde nécessite une réorganisation, voire une rupture, au sein de ces routines articulatoires. La complexité de cette réorganisation est dépendante de la proximité ou de l'éloignement phonologique de la langue maternelle et de la langue seconde. Ce travail se propose d'aborder cette question chez l'apprenant d'une langue seconde. En effet, s'il est établi que l'aptitude des locuteurs non-natifs à acquérir de nouvelles catégories phonémiques est d'autant plus faible que l'acquisition est tardive le débat reste ouvert quant aux mécanismes sous-jacents à ce processus ainsi que le rapport entre la perception et la production. Dans le processus d'acquisition d'une langue maternelle, les capacités perceptives anticipent clairement l'aptitude à produire les sons. Ce travail se propose d'observer si cette chronologie est vérifiée au cours de l'apprentissage d'une langue seconde.

Des locuteurs anglophones (californiens) récemment arrivés en France ont été recrutés pour être enregistrés. Une étude longitudinale, sur deux périodes, a été menée de façon à observer les effets de l'immersion sur l'apprentissage des voyelles. Il s'agit donc de tester des anglophones au cours de leur acquisition du français. Ces anglophones sont testés au bout de trois mois à partir de leur arrivée en France, puis au bout de neuf mois, alors qu'ils suivent une formation accélérée de français. Les tests comportent deux volets. Le premier est un volet perceptif : les apprenants doivent situer les voyelles de leur langue, puis celles du français, sur un espace acoustique (outil développé par René Carré au DDL, UMR 5596, Lyon). Le deuxième est un test de production : les apprenants doivent produire une liste de mots contenant des voyelles critiques dont l'acquisition peut être plus problématique. Les voyelles d'intérêt sont /ɪ/-/y/, /a/ conformément aux résultats obtenus précédemment (Meunier et al., 2005).

Sur la figure 4.e, on observe que la localisation des voyelles /a/ et /ɪ/ est assez proche de ce que font les francophones natifs. En revanche, pour /y/ (absent du système anglais), l'identification est plus tardive mais on observe un effet de la durée d'immersion (plus proche de la cible à 9 mois qu'à 6 mois d'immersion). En revanche, pour la production (figure 4.f), les réalisations des anglophones restent assez éloignées des cibles francophones pour les voyelles /y/ et /ɪ/. Tout

se passe comme si l'espace de réalisation de ces deux voyelles était confondu, même s'il tend vers la distinction de deux espaces après 9 mois.

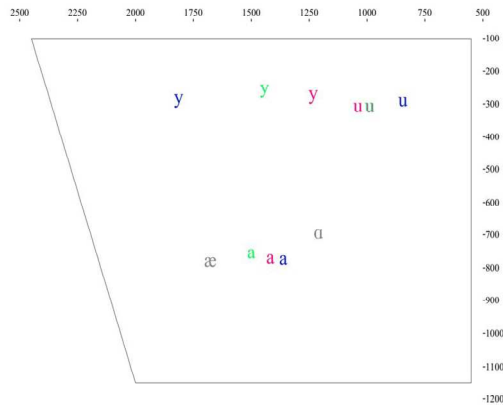


Figure 4.e: Localisation perceptive des voyelles sur le plan F1/F2. En bleu, les voyelles cibles du français (localisées par des francophones); en gris les voyelles correspondantes de l'anglais (localisées par les californiens); en rose, les voyelles cibles du français (localisées par des californiens à 3 mois); en rose, les voyelles cibles du français (localisées par des californiens à 9 mois). Figure extraite de la thèse de P. Péri.

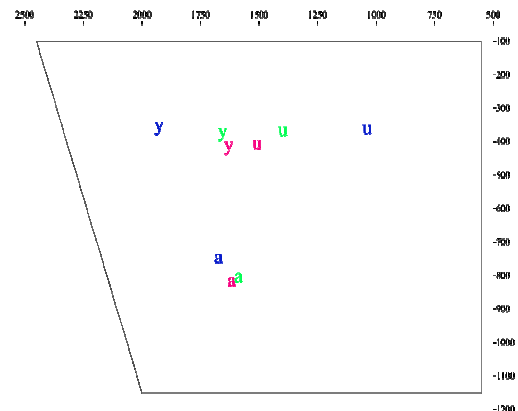


Figure 4.f: Valeurs F1/F2 des réalisations des voyelles. En bleu, les voyelles cibles du français (produites par des francophones); en rose, les voyelles du français (produites par des californiens à 3 mois); en rose, les voyelles du français (produites par des californiens à 9 mois). Figure extraite de la thèse de P. Péri.

En conclusion, la distinction perceptive se met en place à l'issue de l'expérience en immersion. C'est également le cas pour la production mais avec un léger retard: une distinction est opérée en perception sur le contraste /u/-/y/ mais pas de manière suffisamment robuste en production. Ces résultats sont donc en accord avec le modèle de Flege (1995): une différence établie en perception permet de réaliser une distinction articulatoire. Au vu de ces résultats, il semble donc que la perception précède bien la mise en place des distinctions articulatoires.

IV.2 AUX SITUATIONS DE PAROLE

Imaginons un éthologue qui analyserait le comportement des singes au zoo et dans la nature. Ses observations lui permettraient probablement de conclure que le singe se comporte différemment dans les deux milieux. Que devrait-il conclure de ses observations? Que le comportement *réel* des singes est celui que l'on observe dans son milieu naturel ou, à l'inverse, en captivité? En fait, aucune situation n'est plus *réelle* qu'une autre: le comportement

différencié des singes est juste le reflet d'une adaptation à des situations différentes. D'une certaine façon, et c'est sans doute le plus intéressant, les caractéristiques de son adaptation sont révélatrices de sa nature. Et on imagine aisément que différentes espèces présenteront des formes d'adaptation différentes.

Si la comparaison est osée, la capacité qu'ont les êtres humains à adapter leur façon de parler à la situation dans laquelle ils se trouvent n'est peut-être pas si différente. Nous parlons toujours dans une situation donnée, avec un contexte spécifique et dans un état spécifique. Aussi, toute parole est marquée par une somme importante de paramètres qui sont issus d'une part du locuteur lui-même et d'autre part de ce qui lui est extérieur et qui l'entoure, la situation dans lequel il se trouve. Chaque type de situation va donc entraîner un registre de parole adapté à cette situation. On parle alors de *styles* de parole différents ou de phonostyles (Simon et al., 2009).

La notion de *styles de parole* n'est pas simple à aborder sous la forme de rubriques ou de catégories. Il y a, en fait dans son énumération une gradualité qui rend la catégorisation peu satisfaisante. Toutefois, savoir dans quelle situation la parole a été produite est déterminant, tant la nature des productions phonétiques est dépendante de cette situation (Rouas et al., 2010). A ce propos, un débat souvent polémique s'est développé en ce qui concerne la caractérisation des styles de parole, mais également pour ce qui concerne leur légitimité pour l'étude de la parole et du langage oral. Dans son article *In defense of Lab speech*, Xu (2010) entend redonner une légitimité aux travaux portant sur un type de parole très contrôlé. Pourquoi cela est-il nécessaire? Depuis le développement des TGC (Très Grands Corpus) et des outils qui permettent son exploitation de nombreuses recherches se sont tournées vers l'analyse de corpus de parole plus naturelle, réelle source d'observation de phénomènes nouveaux. Le développement de ces travaux a parallèlement conduit à une dépréciation de la parole très contrôlée ou lue. D'une certaine façon, la parole lue et la lecture de séquences isolées, dont est issue une grande partie des travaux phonétiques du XXème siècle serait considérée désormais comme une fausse parole, une parole artificielle uniquement présente dans les travaux des linguistes. La critique principale étant que cette parole n'est en rien représentative des processus mis en œuvre par un locuteur d'une langue donnée lorsqu'il produit de la parole dans les situations les plus courantes. S'agit-il alors de deux systèmes paroles? Ou bien s'agit-il du même système avec des formes différentes ?

Il est difficile d'admettre que nous ayons deux systèmes distincts de production de la parole, l'un pour la parole lue, l'autre pour la parole spontanée. Nous utilisons le même système articulatoire, les mêmes phonèmes, les mêmes mots. La parole lue étant un contexte de

production peu fréquent, il serait peu économique de développer deux systèmes de production distincts, qui supposerait, de fait, deux systèmes de perception distincts.

Il est donc plus raisonnable de penser que nous adaptons notre production à la situation de parole. Toutefois, l'hypothèse sous-jacente des défenseurs de la *Lab Speech* est qu'elle serait une espèce de prototype de la parole spontanée où les propriétés du langage apparaîtraient plus clairement, car dénuée de la très forte variation inhérente à la parole en situation naturelle. Il serait donc juste nécessaire d'observer les caractéristiques de la parole lue pour comprendre les propriétés du langage. Le reste ne serait que variation non interprétable dans le cadre de la linguistique. On retrouve ici la dichotomie très ancienne (langue/parole, Saussure, 1916).

En ce qui me concerne, j'adopte une position différente. Si nous admettons clairement qu'un seul système linguistique régit les différents types de parole, nous faisons l'hypothèse que tout le reste n'est qu'histoire **d'adaptation**, sans privilégier un type de parole comme plus représentatif qu'un autre des propriétés du langage. Rien ne nous permet de penser que les propriétés relevées et analysées sur des corpus de parole lue sont plus fondamentales ou prototypiques que celles de la parole spontanée. Elles sont juste caractéristiques de ce contexte précis, ce qui ne permet pas d'extrapoler sur un autre type de parole. Seule une comparaison de différents types de parole peut nous montrer ce qui est pertinent dans toutes les situations de parole. Notre hypothèse est ainsi qu'une des propriétés fondamentales du langage tient dans cette adaptation entre les situations de parole et dans son fonctionnement.

Par ailleurs, il apparaît que, les distinctions stylistiques propres à un domaine d'étude ne vont pas apparaître avec la même gradualité dans un autre. Par exemple, Joos (1972) propose une distinction en 5 types de styles: intime (*intimate*), familier (*casual*), consultatif (*consultative*), formel (*formal*) et figé (*frozen*). Toutefois, cette catégorisation, plutôt adaptée à la parole spontanée, ne tient pas compte de la parole extrêmement contrôlée qu'est la lecture oralisée qui, néanmoins, est le support d'une grande partie des travaux en phonétique et en psycholinguistique. Il est donc clair que cette gradualité correspond aux travaux portant sur des domaines linguistiques plus vastes que le mot. Les phonéticiens, quant à eux, vont classiquement opposer la parole lue (ou hyperarticulée) à la parole spontanée. De ce fait, ce que les phonéticiens appellent *spontané* correspondra à de nombreuses catégories distinctes dans un autre domaine de recherche. Ce décalage entre les disciplines a fait naître un débat très polémique sur la nature de ce qu'est la parole spontanée. Ce débat, à mes yeux n'a pas lieu d'être, car la définition que l'on donne aux styles de parole est, en fait, dépendante de l'objectif des recherches menées et surtout de l'objet linguistique qui est observé. Si la parole spontanée est décomposable en de nombreux styles bien distincts dans certaines disciplines (prosodie,

syntaxe, discours, interaction, etc.), c'est que cette distinction est pertinente et a du sens pour ces domaines de recherches. En revanche, pour ce qui concerne l'analyse des segments phonétiques, il y a une différence majeure entre la parole lue (contrôlée et préparée) et la parole spontanée (non préparée) (Smiljanic & Bradlow, 2009). D'une certaine façon, une distinction plus fine serait à faire à l'intérieur de la parole lue entre la prononciation d'items isolés (mots, syllabes, logatomes) et la lecture de texte (parole enchaînée) qui fait apparaître des phénomènes de variation déjà très pertinents (Meunier, 2005, 2006). On voit bien que cette distinction, voire même la notion de parole lue, n'a pas de sens dans certaines disciplines (en analyse conversationnelle par exemple), mais elle prend tout son sens pour ceux qui travaillent sur de toutes petites unités linguistiques. En conséquence, lorsque l'on distingue parole lue (préparée) et parole spontanée (non préparée), il est sans doute plus judicieux de parler de *type* de parole plutôt que de *style* de parole.

IV.2.1 Information et variation de la parole

Dans un travail mené avec Philippe Blache, nous avons fait l'hypothèse que la quantité d'information véhiculée dans le message linguistique était une source possible de variabilité de tous les secteurs de l'analyse linguistique.

Contribution:	La répartition de l'information linguistique
Blache, P. & Meunier, C. (2004) "Domaines et propriétés : une description de la répartition de l'information linguistique", Journées d'Etudes Linguistiques, Mai 2004, Nantes, p. 197.	
Blache, P. & Meunier, C. (2004) "Language as a complex system: the case of phonetic variability", VI Congreso de Lingüística General, Mai 2004, Saint-Jacques de Compostelle, Espagne.	

Le processus de la communication parlée suppose qu'un message linguistique est transmis au moyen d'un système caractérisé par un certain nombre de secteurs linguistiques. Ce processus suppose que, pour être compris, un message doit contenir une certaine "quantité d'information". Cette information est, entre autres, contenue dans les divers secteurs du langage (décodage acoustico-phonétique, sémantique, analyse syntaxique, etc.). Chaque secteur possède une structure interne et peut être décrit par un ensemble de propriétés. Or, la transmission du message linguistique est généralement assurée par *l'ensemble* des secteurs. Cela suppose que chaque secteur n'est généralement pas porteur de la totalité de l'information. Il existe alors une relation inversement proportionnelle entre la quantité d'information véhiculée et la nécessité de produire des formes phonétiques prototypiques. Ainsi, lorsqu'un seuil minimum d'information est atteint par quelques secteurs linguistiques, il n'est plus nécessaire que les autres secteurs soient informatifs. Une partie de l'information phonétique peut être sous-spécifiée si les aspects sémantiques, syntaxiques ou prosodiques sont suffisamment

informatifs. Toutefois, cette variation rend complexe l'identification et l'interprétation des indices phonétiques pertinents. Dans la méthodologie de l'analyse phonétique, il est courant de contrôler les situations de production. Réduire la variation consiste ainsi à réduire l'information de façon à ce que le secteur étudié soit le seul secteur garantissant la non-ambiguïté du message. Il va de soi que lorsque l'information linguistique n'est garantie que par un seul secteur, ce secteur doit être totalement informatif pour que le message ne soit pas ambigu (figure 4.g).

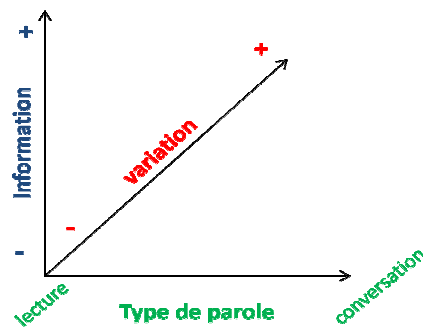


Figure 4.g: Illustration du lien entre quantité d'information contextuelle, type de parole et quantité de variations phonétiques.

Ainsi, l'information linguistique est **dispersée** et **variable**, souvent **incomplète**, et quelquefois instable. C'est pour cette raison que la situation de la parole spontanée est parfois considérée comme une *Adverse Condition* (Mattys et al., 2012), dans la mesure où, effectivement, le secteur phonétique est particulièrement lacunaire si on le compare à une parole très contrôlée. Mais si le secteur phonétique est lacunaire, l'ensemble du message linguistique, lui, ne l'est pas. L'interprétation d'un message linguistique repose donc sur notre capacité à récupérer ces parties d'information et à les mettre en relation. L'ensemble des théories linguistiques cherchent à donner une explication exhaustive du fonctionnement du langage. En particulier, à l'intérieur de chaque domaine linguistique (phonologie, syntaxe, morphologie, pragmatique) il est question de rendre compte systématiquement du fonctionnement des énoncés dans le cadre d'une grammaire. Or, depuis de nombreuses années, les linguistes observent que leur propre domaine de description est souvent insuffisant pour rendre compte, de façon formelle, des énoncés réellement produits par les locuteurs. Il semble donc de plus en plus clair que les différents domaines linguistiques sont en interaction aussi bien en production qu'en perception du langage: la compréhension d'un message linguistique nécessite un ensemble d'informations, éventuellement redondantes, faisant intervenir tous les éléments du langage.

Cette réflexion s'inscrit dans une démarche générale d'élaboration d'un cadre théorique visant à rendre explicite les liens fonctionnels qui interagissent entre les domaines linguistiques.

IV.2.2 La réduction vocalique

La réduction vocalique a été observée en premier par Lindblom (1963) qui constate que les voyelles inaccentuées du suédois se centralisent lorsque le débit augmente. La réduction vocalique (ou centralisation des voyelles) se caractérise par une modification spectrale de l'ensemble des voyelles qui tendent vers un éloignement de la périphérie, donc le centre de l'espace vocalique. Cette modification spectrale est due à un phénomène *d'undershoot* (les cibles articulatoires des voyelles n'étant plus atteintes). Depuis, de nombreux travaux ont pu reproduire cette réduction ou centralisation des systèmes vocaliques qui semble être une constante dans toutes les langues (Gendrot and Adda-Decker, 2007).

Contribution: **Variation des voyelles du français et type de parole**

Meunier C., Floccia, C. (1999) "Syllabe ou mot : quelle unité permet d'identifier les catégories phonétiques?", Actes des 2èmes Journées d'Etudes Linguistiques, "Syllabes", Nantes, pp. 87-92.

Meunier C., Floccia, C. (1997) "La spécificité des types de parole pour la perception de la voyelle", Actes des 1ères Journées d'Etudes Linguistiques, "La voyelle dans tous ses états", Nantes, 158-163.

Jusqu'à quel point la parole de laboratoire est-elle vraiment représentative de la parole naturelle? Dans la parole naturelle, il existe un grand nombre d'informations non présentes dans les mots produits en isolation : les informations prosodiques, syntaxiques, sémantiques et pragmatiques. De ce point de vue, la parole de laboratoire est évidemment plus pauvre que la parole naturelle. Mais d'un autre côté, la parole de laboratoire a des caractéristiques acoustiques et phonétiques nettement meilleures, puisqu'une grande attention est portée à la qualité de l'enregistrement et à la diction des locuteurs. Certaines expériences (Pollack et Pickett, 1969) ont ainsi montré que, coupés de leur contexte, la plupart des mots extraits de discours spontanés ne sont plus identifiables.

Dans cette étude, nous comparons la variabilité des voyelles dans trois modalités de parole: dans des mots extraits de parole continue spontanée, des mots lus en isolation et des mots extraits de phrases lues. Cette dernière modalité est un compromis entre les deux précédents types de parole: il s'agit de parole lue, donc vraisemblablement réalisée plus distinctement que la parole spontanée, mais permettant au locuteur d'utiliser d'autres sources d'information. Les caractéristiques acoustiques et phonétiques des mots extraits de phrases lues devraient se situer entre celles des mots produits en isolation et celles des mots extraits de parole spontanée. Nous avons choisi d'analyser les caractéristiques formantiques et temporelles des voyelles /a/ et /i/. Ces deux voyelles sont acoustiquement et articulatoirement les plus éloignées. Notre étude

permettra peut-être de dégager des caractéristiques de variation propres au type de production de chaque voyelle.

Nous avons sélectionné 30 mots au sein d'un discours spontané (une locutrice conversait librement). Les 30 mots sélectionnés ont ensuite été intégrés dans des phrases porteuses qui ont été lues par la même locutrice (phrases courtes, 7 mots en moyenne, et de structure syntaxique simple). Enfin, ces 30 mots ont été lus isolément toujours par la même locutrice. Le 1^{er} et le 2^{ème} formant de chaque voyelle ont été mesurés manuellement au travers d'une analyse spectrale LPC.

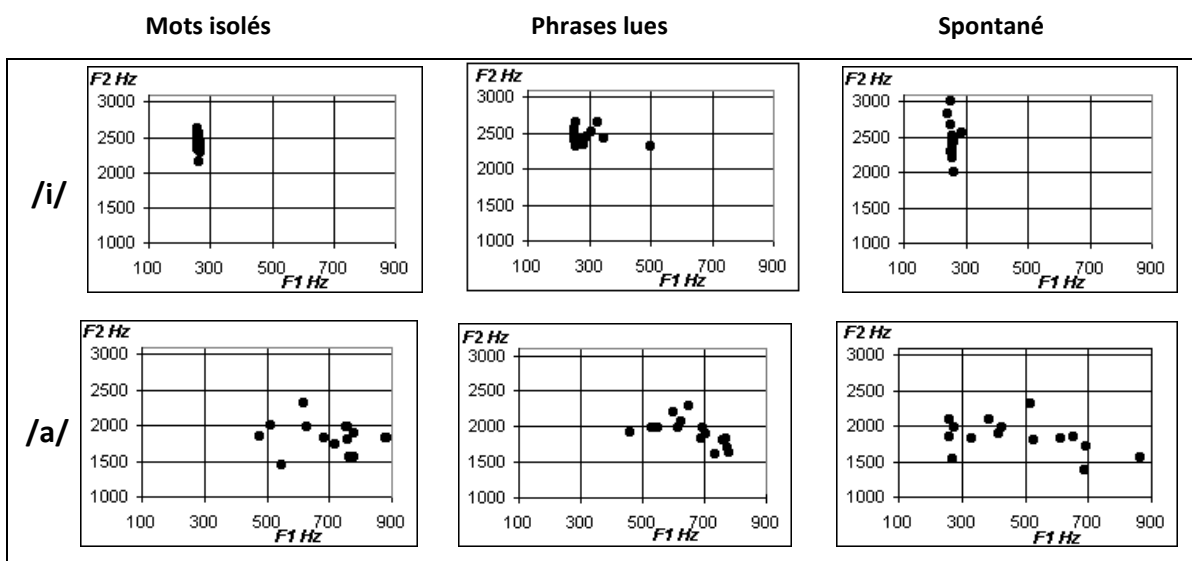


Figure 4.h: répartition spectrale des 15 voyelles /i/ (en haut) et /a/ (en bas) en fonction des trois modalités.

Un premier résultat peu surprenant est que la durée des voyelles est significativement plus importante en condition de mots isolés que dans les deux autres conditions, ce qui signale une diction plus rapide en parole continue qu'en lecture de mots isolés. Un second résultat plus intéressant concerne les dispersions des valeurs de F1 et F2. Tout d'abord les valeurs de F1 et F2 pour /i/ se situent autour des valeurs généralement mises en évidence par les phonéticiens, et ce quelle que soit la modalité de parole (figure 4.h, haut). Ceci signale une grande stabilité de cette voyelle. En revanche, les valeurs F1 et F2 de /a/ varient de manière importante en condition S, un peu moins en condition L, et encore moins en condition I, tout en étant nettement plus dispersées que les valeurs des formants pour le /i/ dans les modalités équivalentes (figure 4.h, bas). Notons que Les valeurs de /a/ en parole spontanée couvrent une partie importante du champ spectral des voyelles du français. Pour expliquer ces effets, des raisons articulatoires peuvent être invoquées: la voyelle /a/ est le phonème qui demande un degré d'aperture maximal, on peut supposer que le /a/ soit éloigné de sa cible articulatoire (et

se rapproche ainsi de l'articulation des phonèmes qui l'entourent) pour des raisons de réduction d'effort articulatoire (Lindblom, 1990). Une autre explication, qui n'exclut pas la précédente, est que la voyelle /a/ est la voyelle la plus fréquente en français. Il est possible que pour l'identifier, il ne soit pas nécessaire de disposer d'indices acoustiques très réguliers et systématiques dans la mesure où sa probabilité d'apparition dans un corpus est très importante.

En résumé, les analyses acoustiques indiquent une grande instabilité des valeurs formantiques du /a/ en fonction de la modalité de parole, avec un maximum de dispersion obtenu en parole spontanée. Ces variations sont-elles perçues par les auditeurs francophones? Pour répondre à cette question, nous avons réalisé une **expérience de transcription (E1)** des voyelles /a/ et /i/ analysées précédemment.

Les 90 voyelles précédemment extraites pour l'analyse acoustique ont été utilisées pour cette expérience. Trois listes ont été constituées. Chaque liste contenait les 30 voyelles d'un type de parole. A ces 30 voyelles ont été ajoutées 15 voyelles non /a/ et non /i/ faisant office de distracteurs. L'ensemble de ces stimuli a été présenté dans un ordre aléatoire fixe.

Toute modalité de parole confondue, les sujets identifient correctement la voyelle /i/ dans 85.8% des cas, contre seulement 26.2% pour la voyelle /a/. Cependant, pour le /a/, 95% des réponses données sont des voyelles proches acoustiquement du /a/ (voyelles mi-ouvertes: /ɛ œ ɔ/). Pour le /a/ comme pour le /i/, le score d'identification correcte des voyelles augmente significativement à mesure que la parole est plus contrôlée (figure 4.i).

Ce travail a pu montrer que si la modalité de parole joue un rôle important dans la variabilité, le type de voyelle est également un facteur déterminant. Les résultats indiquent une variabilité importante à la fois acoustique et perceptive pour la voyelle /a/, et ce d'autant plus importante que la parole est moins contrôlée. Toutefois, on peut s'interroger sur le rôle de cette variabilité: peut-elle gêner - ou aider- la reconnaissance des mots? Nous avons cherché à déterminer si cette variabilité était aléatoire, ou bien si elle était relativement stable à l'intérieur de chaque mot. Des analyses de corrélation post-hoc montrent que pour un même mot, les valeurs des formants de chaque voyelle ne varient pas beaucoup. Ainsi pour le /a/, ces corrélations montrent que les variations sont fonction du mot et/ou du contexte phonétique environnant. De nombreux travaux (Stevens et House, 1963; Schouten et Pols, 1979) ont en effet montré l'impact du contexte phonétique sur les caractéristiques spectrales des voyelles. Cette régularité dans les variations pourrait être utilisée par l'auditeur pour identifier le mot.

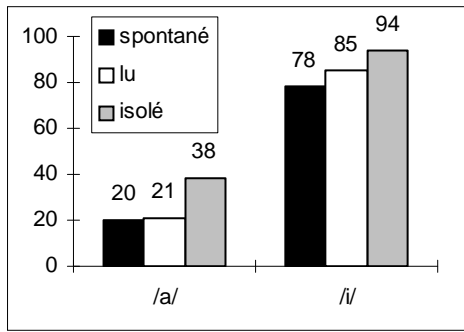


Figure 4.i: Expérience E1, proportion (en pourcentage) d'identification correcte des voyelles /a/ et /i/ en fonction des 3 modalités.

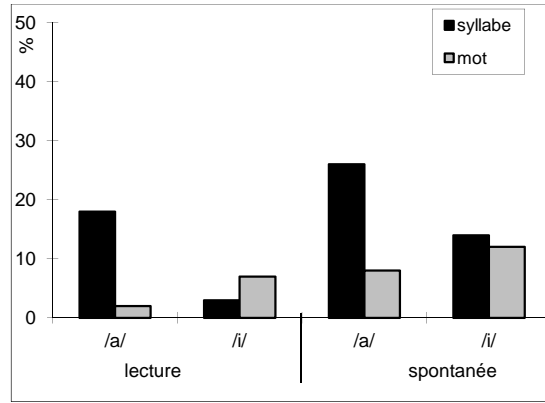


Figure 4.j: Expérience E2, pourcentage de non-détection des voyelles /a/ et /i/ en fonction de la modalité de parole (lecture et parole spontanée) et du contexte (syllabe ou mot).

A partir de ces résultats, nous avons voulu déterminer la nature de cette variation importante de la voyelle /a/ en comparaison avec la stabilité de /i/. Cette variation est-elle aléatoire (la voyelle /a/ est identifiée après la reconnaissance des mots) ou contextuelle (la voyelle /a/ est identifiée grâce au contexte syllabique environnant)?

Pour répondre à ces questions, 22 sujets ont passé une **deuxième expérience (E2)** dans laquelle ils devaient détecter les voyelles /a/ ou /i/ dans deux contextes différents (syllabe ou mot) et selon deux modalités (mots lus et parole spontanée). Les stimuli correspondaient respectivement au mot entier pour la condition *mot* (mots extraits du corpus spontané pour cette modalité) et à la syllabe pour la condition *syllabe* (syllabes extraites des corpus des deux modalités).

Exemples: syllabes: /fa/, /spa/, /ka/, /si/, /vi/, /gri/...
 mots: farine, spaghetti, caler, citron, vitesse, grillé...

Les résultats montrent qu'il y a plus de non-détections en parole spontanée qu'en mots isolés, quelle que soit la voyelle ou le contexte (figure 4.j). On note également que le contexte joue un rôle important pour la voyelle /a/ et bien moindre pour la voyelle /i/. Toutefois, le taux d'erreur maximum enregistré pour la voyelle /a/ en *Syllabe/Parole spontanée* n'est que de 26%. On est très loin des 80% de non identification observés en transcription de voyelles isolées dans Meunier et Floccia (1997).

Au vu des résultats obtenus en production et en perception, il semble que la voyelle /a/ soit nettement plus variable et plus réduite que la voyelle /i/, beaucoup plus stable. Pour les deux voyelles on note un effet de la modalité de parole (E1) pour l'identification de la voyelle, cet

effet étant bien plus marqué pour /a/. On note également un effet du contexte (E2) pour l'identification de la voyelle (essentiellement présent pour /a/): la syllabe augmente l'identification de /a/ et le contexte lexical permet une identification optimale.

Il semble donc que la voyelle /a/ s'éloigne plus d'une réalisation prototypique en parole spontanée que la voyelle /i/, ce qui est confirmé en perception. Ce constat est un peu surprenant dans la mesure où la voyelle /i/, comme toutes les voyelles fermées, est souvent extrêmement réduite dans sa durée et peut être dévoisée ou bruitée. Il ne semble pas que ce type d'altération nuise à son identification. En revanche, la modification majeure de /a/ concerne son timbre, même si sa durée et les autres composantes restent stables. Il semble donc qu'en français, le timbre soit un indice majeur pour l'identification de voyelles.

Enfin, la meilleure identification de /a/ selon la modalité de parole et le contexte suggèrent que cette voyelle est bien plus sensible au contexte que la voyelle /i/. Cette sensibilité pourrait être due à la position ouverte du /a/ qui pourrait s'amenuiser au contact des autres phonèmes, tous plus fermés. En conclusion, on notera que les phonèmes d'une langue ne montrent pas tous la même sensibilité aux phénomènes de réduction.

Contribution: Réduction vocalique: type de parole et systèmes vocaliques

Meunier, C.; Espesser, R.; Frenck-Mestre, C. (2006) "Aspects phonologique et dynamique de la distinctivité au sein des systèmes vocaliques: une étude inter-langue.", Actes des Journées d'Etude sur la Parole (JEP) (26 : 2006 juin 12-16 : Dinard, France), 333-336.

Meunier, C.; Frenck-Mestre, C.; Le Besnerais, M. (2005) "Vowel variability in speech production context: a cross linguistic study.", *Phonetics And Phonology in Iberia* (2005 juin 20-21 : Barcelone, Espagne).

Meunier, C.; Espesser, R.; Frenck-Mestre, C. (2006) "Phonetic variability as a static/dynamic process in speech communication: a cross linguistic study.", *10th Conference on Laboratory Phonology* (2006 juin 29-juillet 1 : Paris).

L'objectif de ces travaux, en lien avec ce qui a été présenté dans la section "s'adapter aux langues", est d'évaluer les mouvements de variation des voyelles selon, d'une part, la propriété de systèmes linguistiques différents et, d'autre part, le type de parole produite. Ces travaux ont été réalisés dans le cadre du projet *Cognitive*.

Lindblom (1963) observe que les voyelles inaccentuées du suédois se centralisent lorsque le débit augmente. La persistance du phénomène de centralisation dans les systèmes vocaliques français et allemand (et non seulement spécifique aux systèmes à accent fort) a conduit Gendrot & Adda-Decker (2004) à conclure qu'il s'agirait autant d'un phénomène purement physiologique (réduction de l'effort articulatoire) que d'un phénomène dû à des contraintes linguistiques. Notre hypothèse est qu'en effet l'argument physiologique ne peut être une explication

exhaustive du phénomène. La centralisation du système entraîne de fait une réduction de la distinctivité globale (moins de distance entre chaque élément). Cette réduction n'est possible que si d'autres aspects de la communication apportent une compensation et donc maintiennent la distinctivité. Cette *variabilité* serait donc *adaptative* (*H & H Theory*, Lindblom, 1990). Mais plus qu'un constat d'une hypo-articulation en parole non contrôlée, il nous semble nécessaire de rendre explicite la dynamique des mouvements de variations.

Trois langues sont étudiées: le français (FR), l'anglais (EN) et l'espagnol (SP). L'anglais distingue entre 13 et 15 voyelles orales alors que le français en distingue entre 10 et 12. En ce sens, la densité des deux systèmes est assez semblable, mais l'organisation du système des deux langues est assez différente (l'anglais possède des indices secondaires tels que la durée et l'accent). L'espagnol comporte un inventaire comparativement moins fourni, avec seulement 5 voyelles, mais dans cet inventaire, nous trouvons les mêmes voyelles qu'en français et en anglais: /a/, /e/, /o/, /i/, /u/. Les trois langues ont été enregistrées dans trois modalités différentes: lectures de voyelles isolées, mots isolés et textes (les mots isolés correspondaient à des mots du texte). Ces trois langues offrent la possibilité de distinguer l'effet de la densité (inventaire plus ou moins fourni), de l'effet de la complexité du système (indices primaires/secondaires). Ce dispositif permet en outre d'observer l'impact du type de parole (quantité de contexte présent) sur la dispersion des voyelles en fonction des caractéristiques des systèmes des trois langues.

Les analyses ont porté sur la mesure du 1^{er} et du 2^{ème} formant de chaque voyelle (détection automatique des formants à l'aide du logiciel *ESPS Entropics* puis vérification manuelle).

Il semble que les changements dus au type de parole vont dans le même sens dans les trois langues : plus la parole est contextualisée, plus l'espace global occupé par chaque système se réduit (centralisation du système en lecture de textes). Ces résultats vont dans le sens d'une hypo-articulation en parole continue. Parallèlement, si l'espace global de chaque système se réduit en parole contextualisée, la dispersion de chaque voyelle, elle, augmente. Par ailleurs, les variations sont sensibles à la structure de chaque système: le système espagnol se réduit considérablement par rapport à celui de l'anglais (figure 4.k).

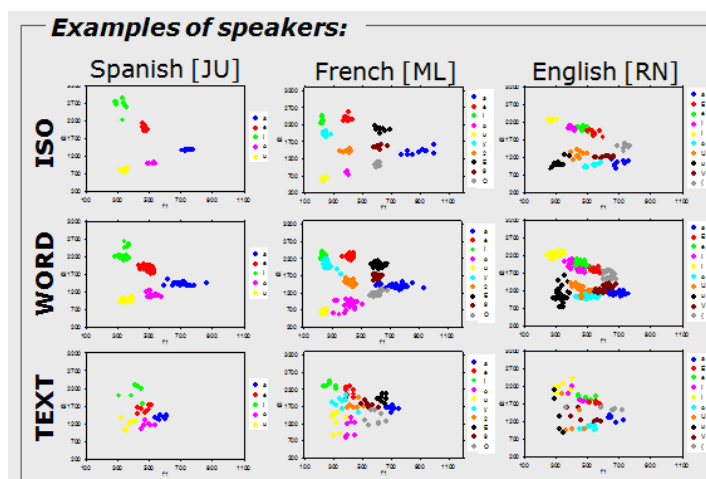


Figure 4.k: Espace vocalique des trois langues pour chacune des trois modalités. Pour chaque langue, exemple d'un locuteur.

Il semble donc que *moins un système est dense, plus il est élastique*. D'une certaine façon, le système de l'anglais est beaucoup moins sensible à la situation de production. Quant au système français il semble intermédiaire, ce qui pourrait être dû à la nature de son système: aussi dense que le système anglais mais moins complexe concernant le type d'indices nécessaires à la distinction des voyelles. Il reste à comprendre si cette hypo-articulation est due à des contraintes physiologiques et/ou à l'information véhiculée par le contexte.

Parallèlement, nous avons observé dans les trois langues que, plus la parole est contextualisée, plus la dispersion de chaque voyelle est importante. Dans les trois langues, mais avec des degrés différents, on passe ainsi d'un système très distinctif à un système peu distinctif (figure 4.l).

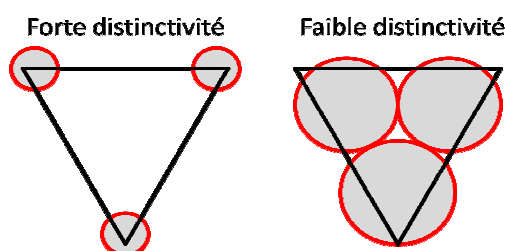


Figure 4.l: représentation de la distinctivité des systèmes vocaliques en fonction de la dispersion de l'espace global et de la dispersion de chaque voyelle

Cette chute de la distinctivité en parole contextualisée peut trouver son explication aussi bien dans une adaptation physiologique (hypo-articulation) que dans une adaptation au type de parole (l'information présente dans le contexte autorise une faible distinctivité phonétique). Ces deux explications ne s'excluent pas, la deuxième pouvant motiver la seconde.

Nous avons calculé un *Indice de Distinctivité (ID)* permettant de visualiser les modifications des systèmes vocaliques en fonction 1/ des propriétés de chaque système; 2/ de la quantité de

contexte. La valeur de cet indice s'obtient en divisant l'Ecart Type des valeurs de l'ensemble du système par l'Ecart Type des valeurs de chaque voyelle, le tout pondéré par le nombre de voyelles du système. Ainsi, plus la valeur de ID est élevée, plus le système est distinctif, et inversement.

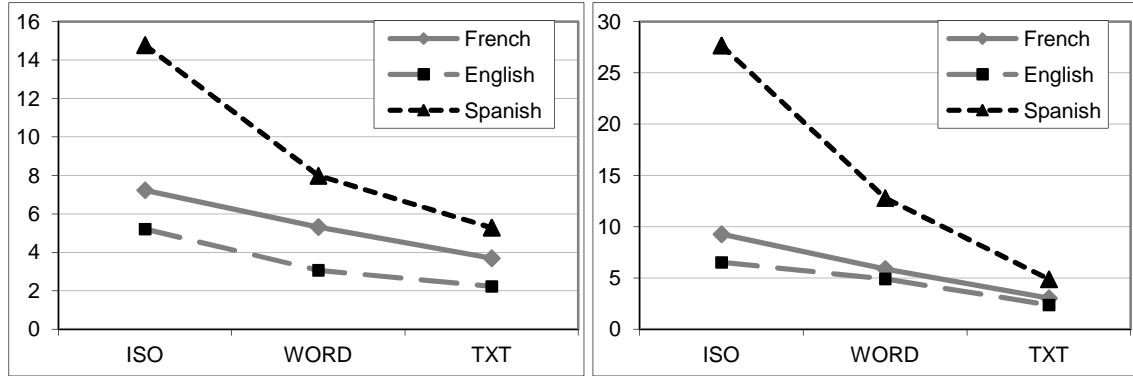


Figure 4.m: Indice de Distinctivité (ID) calculé pour chaque système dans chaque modalité pour F1 (à gauche) et F2 (à droite).

Cet indice rend compte fidèlement des résultats observés: une grande élasticité du système espagnol (surtout sur F2), une plus faible distinctivité des systèmes anglais et français. Et pour les trois langues une distinctivité décroissante au fur et à mesure que la parole est contextualisée (figure 4.m). On remarquera d'ailleurs qu'en parole contextualisée les trois langues tendent vers une même valeur. Cette dernière observation nous semble être une piste très intéressante à explorer: indépendamment de la structure et de la composition des systèmes phonologiques, il est possible de faire l'hypothèse qu'en parole contextualisée, toutes les langues tendent vers un niveau de distinctivité commun ou proche.

Le calcul de cet indice peut nous permettre, par la suite, de comparer systématiquement, des langues, des situations de parole, des locuteurs, des accents, des populations (pathologiques).

Contribution:

Réduction vocalique et propriétés du lexique

Meunier, C., Espesser, R. (2011) "Vowel reduction in conversational speech in French: The role of lexical factors", *Journal of Phonetics*, Volume 39, Issue 3, July 2011, Pages 271-278.

Meunier, C.; Meynadier, Y.; Espesser, R. (2008) "Voyelles brèves en parole conversationnelle". Actes, Journées d'Etude sur la Parole (JEP) (27 : 2008 juin 9-13 : Avignon, France), 97-100.

Dans une première étude, en collaboration avec Y. Meynadier et R. Espesser nous avons cherché à analyser les voyelles très brèves (inférieures à 30 millisecondes) communément écartées par les approches automatiques (Meunier et al., 2008). Ces voyelles brèves apparaissent en très grand nombre dans notre échantillon. Toutes les voyelles du français sont affectées par cette réduction temporelle. Notre étude apporte des observations sur la nature, l'occurrence et la

perception de ces voyelles les plus brèves. Les durées des voyelles en parole conversationnelle sont nettement plus courtes que les durées habituellement observées en parole lue. Mais plus précisément, une très forte concentration des données est observée sur des durées très courtes. Ainsi, après alignement sans aucune correction, 60% des voyelles durent moins de 40 ms. Notre étude porte sur différents aspects en lien supposé avec leur durée brève: 1/ la catégorie lexicale des mots; 2/la longueur syllabique des mots.

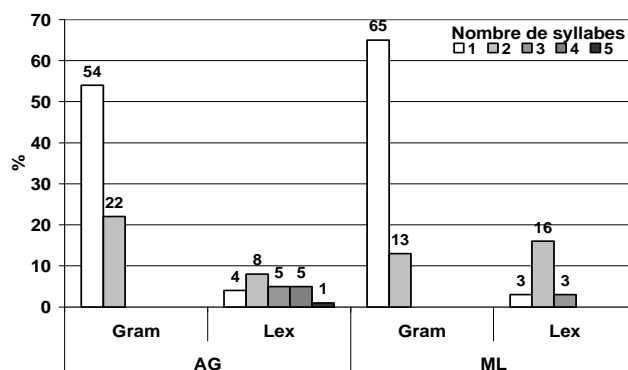


Figure 4.n : Proportion des voyelles extra brèves selon la catégorie lexicale et la longueur syllabique du mot. A gauche le locuteur AG, à droite la locutrice ML.

Ces voyelles extra-brèves se retrouvent très majoritairement dans des mots grammaticaux courts (essentiellement monosyllabiques). Il semble ainsi que les mots grammaticaux soient surreprésentés dans les voyelles extra brèves (figure 4.n). Nous avons en effet noté que dans le corpus CID, la proportion de mots lexicaux (48%) est légèrement plus importante que celle des mots grammaticaux (43%), alors que les voyelles brèves sont trois fois plus fréquentes dans des mots grammaticaux que lexicaux.

Les résultats de cette première étude mettant en lien la durée des voyelles avec les facteurs lexicaux nous ont poussés à mener un travail d'envergure plus vaste dont l'objectif était d'évaluer le rôle de certains facteurs lexicaux connus pour leur influence sur la réduction vocalique.

Nous avons choisi d'observer le rôle de trois facteurs dont la pertinence a pu être mise en lumière dans des travaux antérieurs: 1/ l'allongement final (Jun & Fougeron, 2002), 2/la fréquence lexicale (Pluymakers et al., 2005; Ernestus et al., 2006) et 3/ la catégorie lexicale (mots fonction, mots de contenu) (van Bergem, 1993).

Les paramètres acoustiques analysés étaient la durée des voyelles et les trois premiers formants. L'analyse des durées (table 4.b) montre, bien entendu, des durées bien plus brèves en parole conversationnelle que les durées standard obtenues sur des corpus lus (Bartkova and Sorin,

1987). Ce qui est le plus frappant, en revanche c'est la durée très courte de la voyelle /a/, généralement plus longue que les voyelles fermées. Par ailleurs, la variabilité des durées des voyelles est très importante puisque 50% des voyelles font moins de 60ms quand d'autres font plus de 400ms.

	a	e	ɛ	o	ɔ	ø	œ	ə	i	y	u
CID	65	71		73		84			75	72	86
Bartkova & Sorin (1987)	177	180	175	186	170	186	185	130	170	167	170

Table 4.b: Durées moyennes des voyelles (en ms) dans le CID et durées standard issues de corpus lus (Bartkova & Sorin 1987)

Nous avons donc, ensuite, mis en relation ces durées avec les valeurs spectrales des voyelles. La réduction de l'espace vocalique observée reste cohérente avec les études menées dans ce domaine sur le français et sur d'autres langues (Lindblom, 1963; Adda-Decker et al., 2008). Elle confirme une forte corrélation entre la diminution des durées et la réduction du système, indiquant que la plupart des voyelles réalisées en parole conversationnelle sont produites dans un espace articulatoire très réduit (figure 4.o). Le plus frappant, là encore, est la très forte centralisation de la voyelle /a/ conformément à ce que nous avons déjà observé (Meunier & Floccia, 1997). La réduction est ainsi particulièrement sensible pour les voyelles ouvertes et les voyelles postérieures.

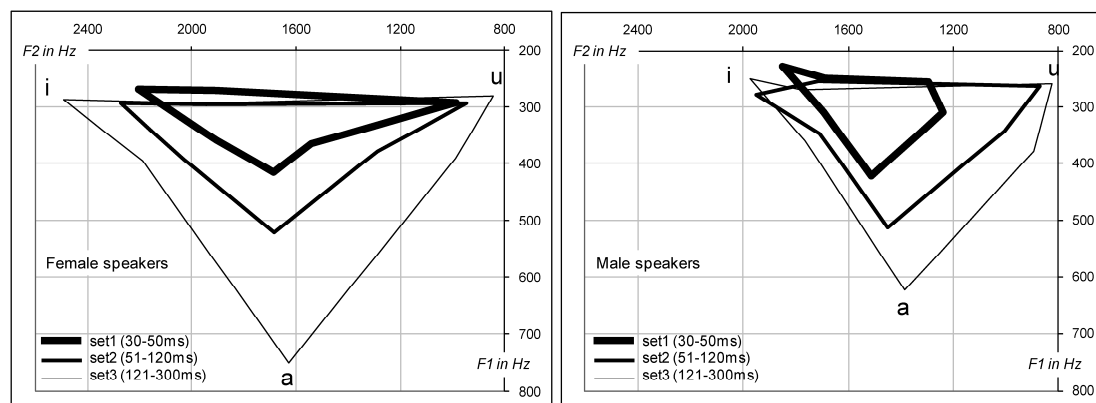


Figure 4.o: Réduction spectrale (en Herz) du système vocalique (femmes à gauche, hommes à droite) du français dans un corpus de parole conversationnelle en fonction de la durée des voyelles (3 groupes de durées différentes).

Un tableau de l'ensemble des voyelles orales du corpus a été construit et permet d'analyser les principaux paramètres acoustiques des voyelles (durées, formants) en fonction de facteurs divers (type de voyelles, structure syllabique, taille du mot en nombre de syllabes, place dans le mot, fréquence lexicale, catégorie morpho-lexicale, etc). Ce tableau nous a permis de relever des

tendances générales concernant les aspects quantitatifs et qualitatifs des réalisations des voyelles en parole conversationnelle.

Nous avons ensuite mis en relation la réduction des voyelles du français avec des facteurs lexicaux (place de la voyelle dans le mot, fréquence lexicale, catégorie lexicale). Nos résultats indiquent un effet marqué de la position syllabique de la voyelle dans le mot: les voyelles non finales sont clairement plus courtes (figure 4.p), ce qui confirme les observations de Jun & Fougeron (2002) sur l'allongement final en français. Nous avons poussé plus loin cette observation pour voir si ces différences temporelles entre le début et la fin du mot étaient également caractérisées par une réduction spectrale. Cette analyse, effectuée sur le F1 de la voyelle /a/ (en raison de sa réduction très marquée) montre le même pattern de réduction que pour les durées (figure 4.q). Toutefois, nos analyses portent de façon indifférenciée sur des syllabes finales accentuées et non accentuées (selon la position de l'*Accental Phrase* AP en français). Il faut donc interpréter ces résultats avec prudence car il est probable que tous les mots pluri-syllabiques ne fassent pas apparaître le même pattern. On peut faire l'hypothèse que les voyelles en position finale d'AP sont bien plus marquées par l'allongement tandis que les voyelles finales de mots mais non finales d'AP pourraient ne pas être allongées.

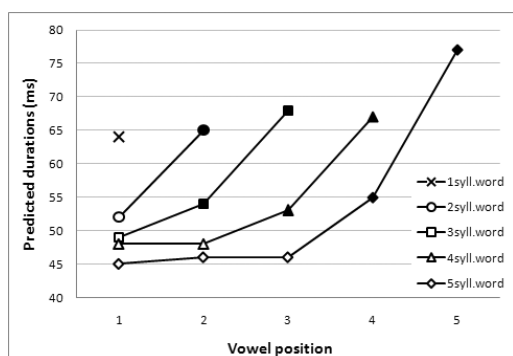


Figure 4.p: Durées des voyelles (en ms) estimées par le modèle. Les durées sont positionnées en fonction de la position dans le mot (axe x) et de la longueur du mot. Les quatre courbes représentent la durée vocalique selon la position de la voyelle dans des mots de 2 à 5 syllabes. Le cercle isolé représente les valeurs pour les voyelles en mots monosyllabiques.

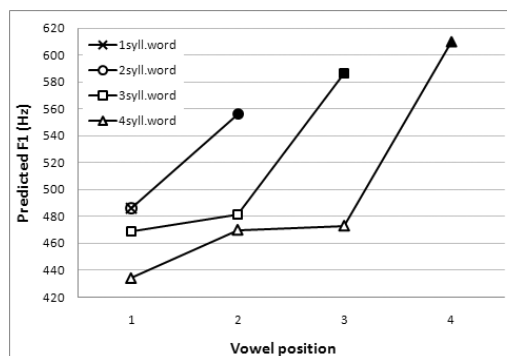


Figure 4.q: Valeurs spectrales du F1 de /a/ (en Hz) estimées par le modèle. Les valeurs sont positionnées en fonction de la position dans le mot (axe x) et de la longueur du mot. Les trois courbes représentent les valeurs de F1 selon la position de la voyelle dans des mots de 2 à 4 syllabes. Les valeurs de /a/ en mots monosyllabiques sont identiques celle de /a/ en première position de mots bisyllabiques).

Des analyses ont ensuite porté sur la réduction des voyelles (toutes les voyelles pour l'analyse des durées et seulement /a/ pour les analyses spectrales) selon la catégorie lexicale des mots porteurs. Nous avons distingué deux catégories:

- les **mots de fonction** (*function words*) contenant les pronoms, prépositions, déterminants et conjonctions
- les **mots de contenu** (*contents words*) contenant les noms, verbes et adverbes.

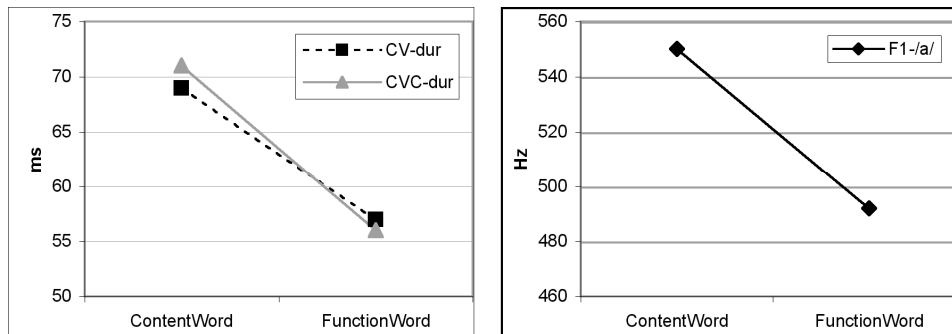


Figure 4.r: Durées (à gauche) des voyelles (en contexte CV et CVC) et valeurs spectrales (à droite) de la voyelle /a/ en fonction de la catégorie morphosyntaxique des mots porteurs.

Un effet marqué de la catégorie des mots est observé (figure 4.r) aussi bien pour les durées de l'ensemble des voyelles que pour ses valeurs spectrales de /a/. Les voyelles réduites se retrouvent plus souvent dans des catégories de mots de fonction. Il y a donc un lien entre la réduction des voyelles et le type de mot produit, ce qui va dans le sens de résultats déjà observés dans des phénomènes de réduction de la parole.

En revanche nous n'avons pas observé d'effet de fréquence des mots indépendamment de la catégorie lexicale. Les mots de haute fréquence étant plus spécifiquement des mots de fonction, ces deux facteurs sont évidemment étroitement corrélés. Il semble que lorsqu'on les distingue, l'effet de fréquence ne ressort plus. Ce résultat est assez surprenant car la fréquence semble être un facteur robuste dans les phénomènes de réduction (Pluymakers et al., 2005; Ernestus et al., 2006).

En conclusion, nous avons pu montrer que la réduction vocalique (aussi bien sous ses aspects temporels que spectraux) était sensible à des facteurs lexicaux. Ces facteurs sont la position dans le mot et la catégorie lexicale. Toutefois, ces deux résultats doivent être confirmés par des études plus poussées. Concernant la position lexicale, il n'est pas certain que le facteur déterminant soit la place de la voyelle dans le mot, mais plutôt place de la syllabe porteuse dans l'*Accentual Phrase*. Pour ce qui est de la catégorie lexicale, je ne suis pas persuadée que la distinction mot de fonction/mot de contenu soit la plus appropriée en parole conversationnelle. Les mots (de fonction essentiellement) sont utilisés dans le cadre de propriétés discursives qui pourraient avoir une influence plus forte sur les phénomènes de réduction phonétique.

Contribution: **Réduction vocalique et voisinage phonologique**

Yao, Y. & Meunier, C. (2014) Effects of phonological neighborhood density on phonetic variation: The curious case of French. Oral presentation at the 14th Laboratory Phonology (LabPhon14) conference. Tokyo, Japan.

Yao, Y. & Meunier, C. (2014). Phonetic variation in French conversational speech. Presentation at the workshop on Corpus-based Approaches to Language Variations at the 2nd Asia Pacific Corpus Linguistics Conference (APCLC 2014). Hong Kong.

Dans ce projet nous avons envisagé de croiser la méthodologie et les questions posées dans deux études différentes. La première étude (Gahl et al., 2012) aborde le rôle du voisinage phonologique sur la réduction vocalique dans un corpus de parole conversationnelle en anglais. La seconde (Meunier et Espesser, 2011) étudie les effets de certaines propriétés phonologiques et grammaticales du lexique sur la réduction vocalique dans la parole conversationnelle en français. Deux facteurs différents ont donc été examinés au travers d'un seul paramètre (la réduction vocalique) et sur un style de parole commun (la parole conversationnelle). L'objectif est donc ici de croiser ces facteurs sur les deux langues étudiées.

L'étude des phénomènes de réduction dans la parole conversationnelle est actuellement en plein essor. Cela implique de comprendre pourquoi et comment les locuteurs d'une langue sont amenés à produire une parole réduite et non canonique, mais aussi comment les auditeurs perçoivent ce type de parole. Il importe également de comprendre la nature des mécanismes de réduction selon les langues. La réduction vocalique dans la parole spontanée est un phénomène très robuste dans de nombreuses langues (Gendrot and Adda-Decker, 2007), mais nous ne savons pas si les facteurs qui induisent cette réduction sont semblables dans toutes les langues. Ce projet vise ainsi à mettre en relation différentes propriétés du lexique avec la réduction vocalique selon la langue.

Ce travail entre dans le cadre des nombreuses études effectuées sur l'effet du voisinage phonologique en perception comme en production de la parole. La nature de cet effet est pour le moins controversé dans la littérature tant les résultats montrent des tendances opposées (l'effet est tantôt inhibiteur, tantôt facilitateur). De plus, il existe peu de travaux en français concernant le voisinage phonologique. Dufour and Frauenfelder (2012) ont pu montrer un effet inhibiteur en perception mais en production les résultats sont moins clairs (Sadat et al., 2014). Nous envisageons donc, pour le Français, la mise au point d'une expérience de production (dénomination d'image) de façon à disposer de résultats comparables pour les deux langues. Cette expérience se tiendra au CEP (LPL) avec des sujets francophones.

Cette première mise en œuvre du projet a consisté en une analyse quantitative de corpus. Pour le français, les 8 heures transcrites du corpus CID (Corpus of Interactionnal Data) ont été utilisées. Le corpus anglais (The Buckeye corpus, Pitt et al., 2005) contenait 40 heures de parole

conversationnelle transcrite en anglais. Conformément à ce qui a été effectué sur le corpus anglais (Gahl et al., 2012) nous avons examiné deux paramètres phonétiques: les variations de durées vocaliques (diminution / allongement) et les variations spectrales des voyelles (réduction / dispersion). Les facteurs étudiés concernent le voisinage phonologique (densité du voisinage, fréquence des voisins et leur interaction). Ces paramètres ont été calculés à partir de la base de données CLEARPOND (Marian et al., 2012). Des mesures lexicales et contextuelles pouvant induire des variations phonétiques (fréquence des mots, probabilité phonotactique, prévisibilité du contexte, débit de parole local, disflunce, etc.) ont été intégrées comme facteurs de contrôle. La construction de deux modèles statistiques (effets mixtes), l'un sur la durée des voyelles, l'autre sur les variations spectrales, a permis d'examiner les effets du voisinage phonologique sur la variation phonétique en français.

Les travaux effectués sur le Buckeye corpus en anglais (Gahl et al., 2012) montraient la relation suivante: plus la densité du voisinage phonologique était élevée, plus la réduction vocalique était importante. Ce résultat allait dans le sens d'un effet facilitateur de la densité du voisinage phonologique en production de la parole (à l'inverse de ce qui est souvent observé en perception). En revanche en français, nos premiers résultats montrent que les mots ayant un voisinage de densité élevée sont réalisés plus longs avec des voyelles plus dispersées (figure 4.s). En revanche, nous n'avons pas trouvé d'effet concernant la fréquence des voisins phonologiques.

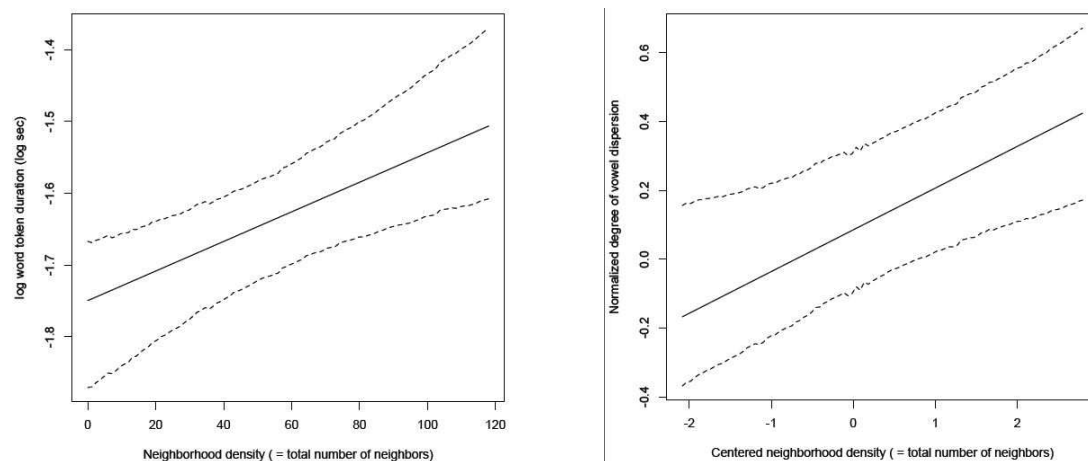


Figure 4.s: Effet du voisinage phonologique sur la durée des mots (à gauche) et sur la dispersion des voyelles (à droite) pour le français.

Ainsi, les résultats actuels tendent à montrer que l'effet du voisinage phonologique sur la variation vocalique en Français est opposé au pattern observé en anglais (Yao & Meunier, 2014). Les mots dont le voisinage phonologique est dense sont hyper-articulés en français mais hypo-

articulés en anglais. Doit-on interpréter cette différence comme un effet inhibiteur dans la production des mots français à forte densité de voisins phonologiques (Sadat et al., 2014)? Ou bien doit-on chercher dans les propriétés des langues des éléments qui pourraient fournir une explication? Pour mieux comprendre cette différence, nous envisageons d'approfondir de nombreux aspects peu connus en français (mais mieux renseignés en anglais) concernant l'effet de certaines propriétés linguistiques (structure morphologique, lexicale, prosodique, discursive, etc.) sur les variations phonétiques. Nous faisons en effet l'hypothèse que ces propriétés (différentes selon les langues) pourraient influencer les phénomènes de réduction et, en conséquence, apporter une explication sur les différences observées entre le Français et l'Anglais concernant le voisinage phonologique.

IV.3 AUX CONTRAINTES

Chaque individu dispose, a priori, de l'équipement physiologique nécessaire à la production de toutes les langues du monde. Cet équipement lui permet de produire de la parole dans des conditions optimales et de se faire comprendre dans de bonne condition également. Cet équipement comprend lui-même des contraintes qui agissent sur sa production: plus son débit va s'accélérer, moins il sera capable d'atteindre des cibles articulatoires caractéristiques de l'hyper-articulation. Ces contraintes physiologiques conditionnent en partie les caractéristiques de son adaptation.

Toutefois, en plus de ces contraintes inhérentes au système, il arrive que le corps fasse défaut. Différentes pathologies entravent ainsi la production du langage. Cette déficience peut affecter différentes composantes du langage, comme le lexique (aphasie), la syntaxe, le système de référence (schizophrénie) ou encore la production de la parole (dysarthries). Les mécanismes d'adaptation caractéristiques de la parole saine sont alors perturbés. Ce sont ces perturbations que nous allons aborder dans cette partie.

IV.3.1 Les populations pathologiques

J'ai souhaité dire quelques mots sur la notion de population car dans la plupart de nos recherches sur la parole une forme de lissage est souvent effectuée de façon à réduire la variabilité interlocuteur et à conserver une population homogène. Cela semble assez logique car c'est là un facteur de variation que l'on ne souhaite pas traiter. Une grande partie des travaux en

phonétique et en psycholinguistique est ainsi fondée sur une population jeune (20 à 30 ans) issue d'un milieu socio-éducatif assez homogène, je veux parler des étudiants des universités.

Mes travaux sur des populations atteintes d'une pathologie m'a permis d'accepter que cette condition de population homogène est parfois une utopie. Comme la plupart de mes collègues, j'ai souvent regardé avec beaucoup de perplexité les travaux portant sur des observations cliniques de patients. Je comprenais mal pourquoi il s'agissait d'études de cas, souvent très limitées ne permettant pas de généraliser l'observation. D'une certaine façon, je le comprends désormais beaucoup mieux et c'est avec une certaine humilité que j'aborde les travaux de mes collègues cliniciens.

Dans la plupart de nos travaux, les locuteurs/auditeurs sont *choisis* en fonction de leurs caractéristiques. C'est, somme toute, logique puisque, justement, nous avons ce choix. Pour les populations pathologiques, les locuteurs sont représentatifs de la population globale, excepté pour ce qui concerne l'épidémiologie de la maladie (âge pour les dysarthries, par exemple). De ce fait, les populations pathologiques étudiées présentent une très forte hétérogénéité dans leur production. En premier lieu parce que des symptômes différents peuvent caractériser une même pathologie (débit rapide et saccadé, ou bien très lent chez les dysarthriques). Ensuite parce que l'environnement socio-éducatif peut être extrêmement varié, ce qui a de fortes conséquences aussi bien pour la lecture oralisée que pour la production spontanée. D'autres facteurs personnels peuvent également avoir une grande influence sur la parole produite : l'attitude face à la maladie (le patient peut être soit combatif soit découragé et déprimé); le rapport avec l'interlocuteur clinicien/expérimentateur (distance, confiance, méfiance, soumission). Mais le paramètre probablement le plus complexe à appréhender est la **stratégie de compensation** qu'utilisent (ou non) les patients pour conserver leur intelligibilité, mais surtout leur communication avec autrui. Cette stratégie, en elle-même, est un **mécanisme d'adaptation** et, en cela, tout à fait caractéristique d'un fonctionnement linguistique. Sa complexité tient dans le fait que chaque patient peut utiliser des stratégies différentes (maintenir le débit ou la prosodie au détriment de l'articulation, ou inversement, faire des pauses, parler peu, choisir ses mots, etc.).

Ainsi, la difficulté principale lorsque l'on décrit les caractéristiques phonétiques d'une pathologie du langage réside dans ce fait : on ne décrit pas une maladie, on décrit des locuteurs atteints d'une maladie et utilisant des mécanismes d'adaptation spécifiques dans l'objectif de maintenir la communication.

IV.3.2 La parole dysarthrique

L'étude de la parole dysarthrique dans mon parcours est venue par le biais d'un premier projet ANR dirigé par Cécile Fougeron (DesPhoAPaDy). Si l'étude de la dysarthrie était déjà présente au LPL depuis longtemps (Création de la base Aix-Park, travaux de Danielle Duez), je ne m'y étais pas penchée jusqu'à ce que C. Fougeron me demande de participer à ce projet. J'y ai trouvé un cadre supplémentaire permettant de comprendre le fonctionnement de la variabilité et ses limites. L'objectif de ce premier projet était essentiellement de constituer une base de données des différents types de dysarthries à partir de corpus dont certains étaient déjà existants et d'autres étaient à constituer (Fougeron et al., 2010).

La dysarthrie regroupe un ensemble de maladies dont la caractéristique commune est une atteinte du système nerveux central ou périphérique dégradant la transmission des messages neuraux vers les muscles sollicités dans la production de la parole (Darley et al., 1975). Cette atteinte entraîne ainsi un déficit dans l'exécution des mouvements nécessaires à la production de la parole. L'ensemble des composantes phonatoires impliquées dans la parole (respiration, phonation, articulation et résonance) sont affectées. Plusieurs maladies, plus ou moins fréquentes ou connues, sont référencées sous cette atteinte¹:

- La **maladie de Parkinson** qui est une affection du Système nerveux central entraînant un déficit en dopamine
- La **sclérose latérale amyotrophique** (SLA), qui est une maladie neuro-dégénérative (la plus fréquente et la plus grave maladie du motoneurone)
- Les **ataxies cérébelleuses** qui sont la conséquence d'une lésion directe du cervelet ou de ses voies efférentes ou afférentes
- Mais aussi, sclérose en plaques (SEP), accidents vasculaires cérébraux, traumatismes crâniens, etc.

Evidemment, la manifestation clinique de la dysarthrie dépende de plusieurs facteurs dont la maladie qui y est associée, mais aussi, le degré de sévérité du trouble, l'âge dans la maladie ou encore l'état psychologique du locuteur. Pour décrire cette pathologie les cliniciens évoquent le plus souvent les caractéristiques suivantes: faiblesse, lenteur, incoordination, réduction de l'amplitude des mouvements, altération du tonus musculaire ou encore rigidité, spasmes, etc.

Les conséquences de ces altérations sur le plan phonétique se retrouvent sur les trois composantes de la production de la parole : dysfonctionnement respiratoire (variation du débit,

¹ Pour une description exhaustive de la dysarthrie, voir l'ouvrage d'Auzou et coll., Les dysarthries (Auzou et al., 2007)

hypotonie), dysfonctionnement laryngé (monotonie et irrégularité prosodique, perte du contrôle de la hauteur de la voix), dysfonctionnement articuloire (imprécision des consonnes, dispersion et centralisation des voyelles, nasalisation). Les descriptions phonétiques de la dysarthrie restent assez rares et sont souvent issues de bilans perceptifs plus ou moins standardisés (Darley et al., 1969). Concernant le français, les travaux de D. Duez font référence. Dans un travail portant sur la description acoustique des occlusives produites par des locuteurs parkinsoniens (en comparaison avec des locuteurs contrôles), Duez (2007) note les caractéristiques suivantes: diminution des durées, absence fréquente du relâchement explosif de l'occlusive, présence de formants (donc réalisation approximante), présence de bruit sur la tenue de l'occlusive (occlusion non complète), dévoisement plus fréquent. Pour plusieurs de ces réalisations atypiques, Duez note qu'elles sont également présentes chez les locuteurs contrôles, mais moins fréquemment. Ces caractéristiques ont également été retrouvées dans d'autres langues (Logemann & Fischer, 1981; Weismar, 1984; Ackerman & Ziegler, 1991). Toutefois, plus récemment dans une étude portant sur trois populations dysarthriques (Parkinson, SLA et Ataxie cérébelleuse) Kocjančič Antolík & Fougeron (2013) notent que si la déviance des consonnes est plus importantes chez les dysarthriques que chez les locuteurs sains, seulement 8% des consonnes sont considérées comme déviantes chez les pathologiques.

Le point de départ important, concernant mon intérêt pour l'analyse phonétique de cette pathologie, est que, comme l'a mentionné D. Duez, les formes phonétiques de la déviance en dysarthrie sont semblables à celles que l'on peut rencontrer dans la parole du locuteur sain dans des types de parole peu contrôlés: les occlusives bruitées, le dévoisement des consonnes voisées (et inversement), les réalisations approximantes, etc.

D'une certaine façon, l'atteinte motrice dont souffrent les populations dysarthriques entraîne une hypoarticulation dont les conséquences phonétiques pourraient être semblables à celles que l'on observe en parole relâchée. Autrement dit, le locuteur sain évoluerait dans le cadre d'une **hypoarticulation adaptative**, tandis que le locuteur dysarthrique serait caractérisé par une **hypoarticulation contrainte**.

C'est sur ces réflexions que le projet ANR TYPALOC (Typologie, Adaptation et Localisation) dont je suis responsable) a été développé: puisque les déviances observées chez les locuteurs dysarthriques sont semblables aux variations et réductions observées chez le locuteur sain, pourquoi les dysarthriques sont-ils perçus comme locuteurs pathologiques? Si le caractère pathologique ne tient pas dans la nature des déviances, il faut le chercher ailleurs. Deux hypothèses complémentaires sont proposées:

1. **L'adaptation** au style de parole est une caractéristique du locuteur sain dans sa propre langue; le trait pathologique pourrait tenir dans l'incapacité qu'ont les locuteurs à respecter cette adaptation au style de parole.
2. Les formes de réduction ou de variation chez le locuteur sain ne sont pas localisées aléatoirement. Leur distribution est organisée par le système linguistique. Chez le locuteur dysarthrique, la **localisation** des déviations ne serait pas fonction du système linguistique, mais de la déficience du système moteur¹.

Ces deux aspects ont été abordés au cours de deux premiers travaux qui sont développés ci-dessous.

Contribution: **Dysarthrie et adaptation au type de parole**

Fougeron, C., Audibert, N., **Meunier**, C., Kocjančič Antolík T. (2014) "Variation in the production of French vowels: physiological constraints and communicative demands", International Workshop of Language Production (2014, July 16-18, Geneva, Switzerland)

Fougeron, C., Audibert, N., **Meunier**, C. (2013), " La réduction vocalique en parole dysarthrique est-elle sensible au style de parole ?", 5ème Journées de Phonétique Clinique, Liège, Belgique, octobre 2013.

En parole saine, la réduction vocalique est fortement liée au débit (Gendrot and Adda-Decker, 2005; Meunier and Espesser, 2011) et au style de parole (Meunier et al., 2005). Plus le style de parole est relâché et les voyelles courtes (parole spontanée), moins les cibles vocaliques sont atteintes. Une réduction vocalique est aussi observée en parole dysarthrique (PD) lue. Elle apparaît sous forme de réduction spectrale ou d'instabilité des cibles vocaliques, liées ou non à des facteurs temporels, en fonction des types de dysarthries (Turner & Tjaden, 2000; Fougeron & Audibert, 2011). Quelle est alors l'incidence du style de parole sur la réalisation des voyelles en PD ? La PD spontanée est-elle (a) comparable à la lecture où la réduction vocalique serait déjà maximale ? (b) plus réduite qu'en lecture sous l'effet des mêmes contraintes que la parole saine? (c) moins réduite qu'en lecture où la production est fortement contrainte ? D'un point de vue perceptif, nos observations suggèrent une grande variabilité indépendante des pathologies : certains locuteurs sont perçus plus intelligibles en parole spontanée qu'en lecture tandis que pour d'autres, l'inverse est observé.

Cinq locuteurs atteints de SLA (Sclérose Latérale Amyotrophique) ont été enregistrés durant la lecture d'un texte (Le cordonnier) et en parole spontanée (narration de leur journée ou d'un événement de leur vie). Quatre types de voyelles ont été extraits de ces deux types de corpus: /i,

¹ Nous verrons que cette hypothèse, sans être invalidée, a pu être réorientée après l'étude portant sur la liaison.

e, a, o/. Pour chaque exemplaire de voyelles, nous avons pu recueillir de 4 à 12 exemplaires par locuteur.

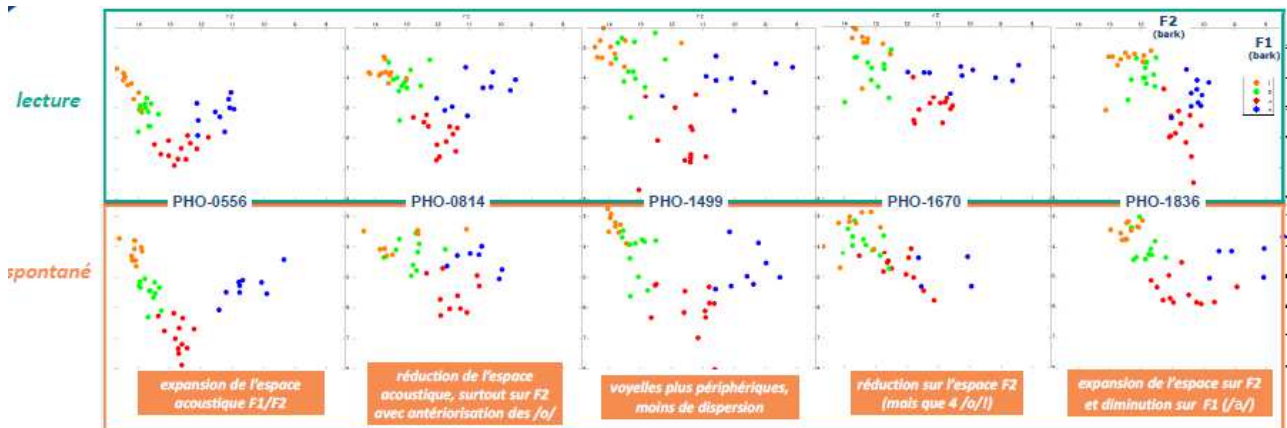


Figure 4.t: représentation des triangles vocaux des cinq patients dans l'espace F1/F2 dans les deux conditions de parole: lecture du texte (en haut) et parole spontanée (en bas).

Contrairement aux variations décrites chez les locuteurs sains en fonction du style, en parole spontanée (vs. lecture) nos 5 locuteurs dysarthriques ne montrent :

- pas de réduction systématique de l'espace vocalique (2 locuteurs réduisent, 2 augmentent, 1 reste stable)
- pas de compression de l'espace de variation sur F1
- pas de réduction de durée vocalique
- pas de corrélation entre durée et centralisation

Ces résultats (figure 4.t) traduisent un comportement différent des locuteurs SLA comparé aux locuteurs sains. Ils ne semblent pas adapter leur production au type de parole produite. D'une certaine façon, la réduction de l'espace est déjà présente en condition de lecture si l'on compare leur productions à celle des sujets sains (Baseline: 26 locuteurs sains enregistrés en lecture) (figure 4.u).

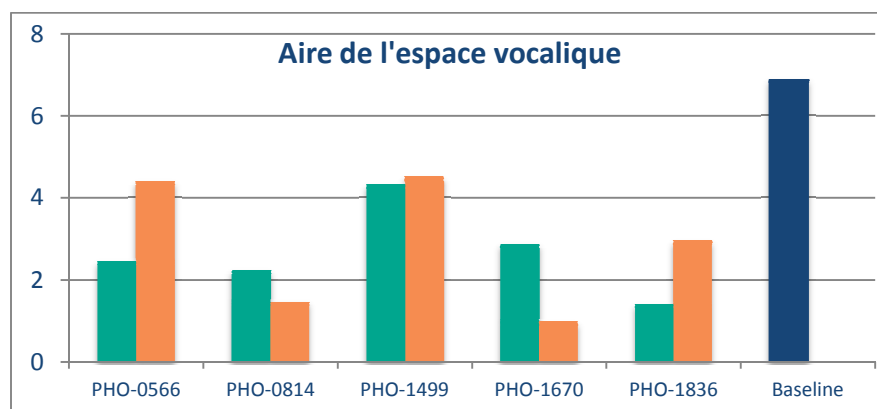


Figure 4.u: Calcul de l'aire de l'espace vocalique des cinq patients en fonction du type de parole: lecture de texte (en vert), parole spontanée (en orange).

Ces premiers résultats semblent aller dans le sens de notre première hypothèse concernant la non adaptation des locuteurs dysarthriques au type de parole qu'ils ont à produire. Toutefois, cela demande à être confirmé avec un nombre plus important de locuteurs et une comparaison systématique avec des locuteurs contrôles dans les mêmes situations de parole.

Contribution: **Dysarthrie et stratégies linguistiques**

Meunier, C., Legou, T. Vidil, C. & Ortega, M. (2013), "Elision des consonnes obstruantes dans la maladie de Parkinson: déficit moteur ou propriété linguistique?", 5ème Journées de Phonétique Clinique, Liège, Belgique, octobre 2013.

L'affaiblissement des consonnes est l'une des caractéristiques majeures de la parole chez les patients atteints de la maladie de Parkinson (Ackermann H. & Ziegler W., 1991). Ce phénomène est également présent chez le sujet sain dans des styles de parole peu contrôlée (Duez, 2003) mais n'est pas perçu comme une altération. Une question centrale est de savoir si la réduction articulatoire présente en parole dysarthrique est uniquement dépendante du déficit moteur, ou bien si, comme pour le sujet sain, elle est en partie intégrée dans un fonctionnement linguistique.

Pour répondre à cette question, nous avons comparé l'élision des consonnes obstruantes au cours de la lecture d'un texte (La chèvre de monsieur Seguin) chez des **Locuteurs Parkinsoniens (LP)** et chez des **Locuteurs Sains (LS)** dans des positions phonologiques propices à l'élision (en position de liaison et au contact du schwa).

114 consonnes du texte ont été retenues. Il s'agissait des occurrences des six obstruantes suivantes: /p, b, t, d, s, z/. 11 étaient en position de liaison (7 facultatives et 4 obligatoires), 16 précédaient un schwa, et les 89 se trouvaient en contexte ne favorisant pas l'élision. 15 LP issus de la Base de Données Aix-Parkinson (Ghio et al., 2012) ont été sélectionnés en fonction de leur degré de sévérité (item parole UPDRS de 1 à 3). 10 LS ont également été extraits de la base pour les analyses (très souvent les conjoints des patients enregistrés lors des visites).

Au vu des résultats, les LP omettent plus de consonnes de liaisons que les LS (48% contre 34%, toutes liaisons confondues, figure 4.v). Mais ce supplément d'élisions est essentiellement le fait des liaisons facultatives. Concernant les liaisons obligatoires on n'observe pas de différences entre les deux populations. Il semble donc que les LP tendent à produire moins de liaisons lorsqu'ils ont le choix de le faire.

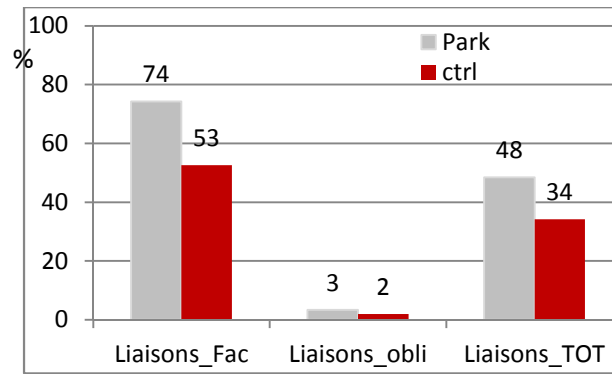


Figure 4.v: Pourcentage d'élision des liaisons pour les locuteurs parkinsoniens et les locuteurs contrôles

Ces résultats sont à confirmer avec un nombre plus important de locuteurs (dans les deux populations). En effet, la liaison est un phénomène complexe à analyser et sa réalisation dépend de facteurs parfois très difficiles à contrôler (origine géographique, profil socio-éducatif, expérience de lecture, etc.). Ces facteurs sont évidemment encore plus difficiles à contrôler chez les patients (souvent âgés) puisqu'ils ne sont pas sélectionnés, a priori, en fonction de cela.

En revanche, les résultats concernant les consonnes en contexte schwa sont moins dépendants du profil du locuteur. Et, là encore, on observe que les LP omettent plus de consonnes que les LS. Le plus intéressant est que, en dehors de ces deux contextes, l'élision des obstruantes ne diffère pas entre les deux populations (figure 4.w). Autrement dit, le LP semble concentrer les élisions des obstruantes dans des contextes qui favorisent cette élision également chez les LS.

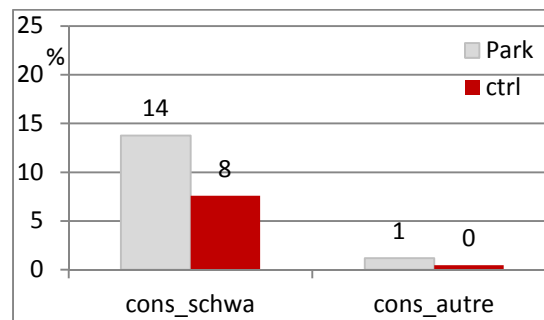


Figure 4.w: Pourcentage d'élision des consonnes en contexte schwa et dans d'autres contextes pour les locuteurs Parkinsoniens et les locuteurs contrôles

Ces résultats montrent que 1/ les consonnes sont plus souvent élidées chez les LP; 2/ ces élisions se font dans le cadre d'une position linguistique favorisante. Comme tous les locuteurs qui cherchent à préserver leur intelligibilité, les LP semblent utiliser une stratégie de préservation des éléments linguistiques primordiaux. Les élisions et réductions affectent ainsi plus favorablement des positions propices à la réduction. Ces stratégies sont par ailleurs celle utilisées par les sujets sains en parole spontanée (Adda-Decker et al., 2012).

Ces résultats semblent contredire en partie notre hypothèse concernant la localisation des déviances chez les dysarthriques: "*Chez le locuteur dysarthrique, la **localisation** des déviances ne serait pas fonction du système linguistique, mais de la déficience du système moteur.*" Il semble toutefois que les dysarthriques parviennent à conserver des stratégies linguistiques de réduction malgré leur déficit moteur. Les contraintes dues au déficit moteur seraient ainsi intégrées au sein de contraintes linguistiques.

IV.4 BILAN

Nous avons développé dans cette partie un ensemble de travaux et de propositions qui vont dans le sens d'une vision du langage comme un système symbolique (composé d'unités abstraites et organisées) actualisé sous la forme de mécanismes d'adaptation propres à ce système. Cette adaptation peut influencer le système lui-même et, à ce titre, fait partie intégrante du fonctionnement linguistique.

L'adaptation dans le processus de production est une caractéristique fondamentale de la parole. Son fonctionnement est donc dépendant :

- des composantes de chaque langue (phonotactique, lexicale, syntaxe, prosodie, structure du discours, etc.) et de leurs propriétés respectives
- des caractéristiques du locuteur (âge, sexe, accent, etc.)
- de l'information véhiculée dans une situation de communication des contraintes physiologiques (de base, et dues à une pathologie)

Les formes de variations et **réductions** présentes en parole spontanée représentent ainsi des **conséquences du mécanisme d'adaptation** tout comme les **productions hyper-articulées** issues de la parole lue. Analyser les formes de réduction dans la parole, c'est ainsi rendre explicite une partie des interactions entre les secteurs linguistiques. Dans cette optique, j'envisage donc de poursuivre mes investigations concernant les formes et les causes des phénomènes de réduction (voir *Perspectives* à la fin de ce document).

V. PERSPECTIVES: RELATIONS ENTRE PROPRIETES DES SYSTEMES LINGUISTIQUES ET MECANISMES D'ADAPTATION

Les perspectives de travaux mentionnées ci-dessous sont issues des bilans effectués en fin de chaque partie. Je ne développerai pas les projets en cours (projet TYPALOC, projet VARIAMU) dont la poursuite est une finalisation des contributions qui ont été présentées. Seront donc développés ci-dessous des travaux qui me semblent pertinents quant aux questions soulevées précédemment.

V.1 LES SONS DES LANGUES ET LEUR USAGE

Systemes phonologiques, phonotactique et usage

Dans ce projet, il est question de mener des études portant, non pas sur la réalisation phonétique des sons, mais sur leur distribution dans différentes langues. Les langues en question pourraient être, dans un premier temps, des langues bien documentées pour lesquelles on dispose de descriptions phonétiques et phonologiques fournies ainsi que de corpus de parole spontanée annotés (anglais, danois, allemand, néerlandais, espagnol). Mes collaborations avec des collègues de Hong Kong permettent également d'envisager des travaux sur le chinois mandarin ou le taïwanais.

Le recueil d'informations (déjà disponible pour plusieurs de ces langues) porterait sur l'inventaire du système phonologique (voyelles et consonnes), le système prosodique, le système phonotactique. L'existence de corpus de parole spontanée annotés pour ces langues permet d'obtenir des fréquences d'usage des phonèmes et des mots (sous la forme des travaux menés dans Meunier, 2012).

Le recueil de ces données permettra ensuite de mener des travaux portant sur la relation entre certaines formes de réduction et la nature des systèmes linguistiques.

Usage des gestes articulatoires dans les langues

Ce travail fait suite à l'étude menée sur la comparaison de la fréquence des gestes articulatoires dans l'inventaire phonologique et dans la parole spontanée (Meunier & Espesser, 2011). Pour réaliser cette étude, il est juste besoin de connaître l'inventaire phonologique des langues (voir ci-dessus), puis de faire correspondre cet inventaire avec les gestes articulatoires déclinés dans

la phonologie articulatoire (Browman and Goldstein, 1992). Comme dans notre étude de 2011, il faut ensuite pondérer le poids de chaque geste en fonction de la fréquence des phonèmes en parole spontanée dans chacune des langues. On pourra ainsi obtenir une fréquence d'utilisation (théorique) des gestes dans différentes langues et envisager un écart possible (ou un équilibre) par rapport à la fréquence des gestes dans l'inventaire phonologique de la langue. En effet, si certains phonèmes sont présents dans toutes les langues, d'autres sont plus rares et requièrent des gestes moins usités (articulation pharyngale, protrusion, etc.). Les résultats permettraient d'observer 1/ si les gestes plus rares s'effacent dans l'usage, 2/ si certains gestes sont sur-représentés, 3/ si toutes les langues tendent vers une utilisation similaire des gestes articulatoires.

Fréquence des enchaînements articulatoires dans les langues

Toutefois, si la fréquence des gestes dans différentes langues est une information importante, elle ne nous donne pas une information très complète sur les enchaînements articulatoires ni sur son implication concernant les phénomènes de réduction. Or, à partir du moment où l'on dispose d'une correspondance entre phonèmes et gestes articulatoires, il est possible d'éditer une annotation des corpus de parole en gestes articulatoires attendus (et non ceux qui sont effectivement réalisés). Chaque phonème pourrait donc recevoir un niveau d'annotation pour chacune des variables (Lip Protrusion, Lip Aperture, Velum, Glottis, Tongue Tip Constriction Degree, etc.) elles mêmes spécifiées par les dimensions appropriées (VEL: closed, LA: wide, etc.) (figure 5.a).

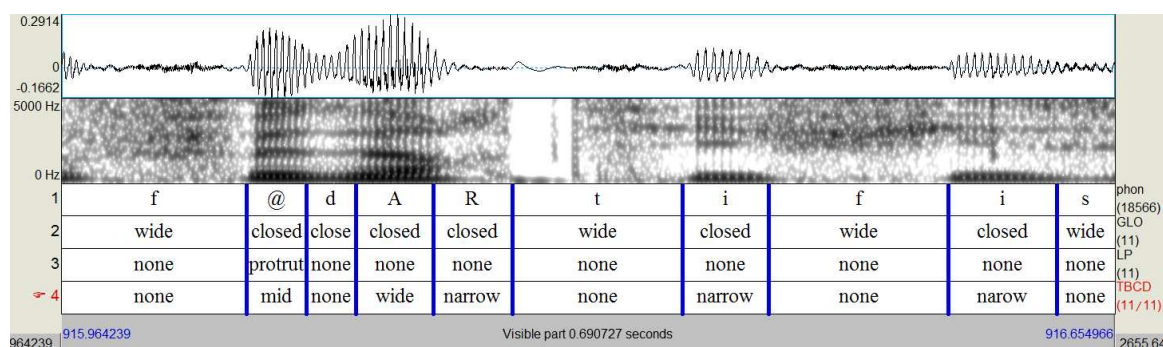


Figure 5.a: Séquence *Feu d'artifice*. Exemple d'annotation possible des variables (GLOttis, Lip Protrusion et Tongue Body Constriction Degree). Pour chaque variable les dimensions appropriées sont déclinées. L'annotation *none* signifie que la variable n'est pas sollicitée pour l'articulation de ce phonème.

J'envisage donc de produire cette annotation en collaboration avec Brigitte Bigi grâce au logiciel Spass (Bigi, 2012). Ces données permettront d'obtenir deux types d'information.

1. La fréquence des enchaînements articulatoires sollicités, et donc des hypothèses sur la localisation de la coarticulation et des réductions (lorsque les positions sont distantes).

2. Des données sur les gestes présentant une variation, par une comparaison entre les dimensions attendues (annotées) et la réalisation acoustique effective (obtenues grâce à des analyses automatiques quand cela est possible).

V.2 INTERPRÉTATION DES FORMES DE RÉDUCTION

Investigations articulatoires

Je l'ai précisé dans le bilan de la partie "Parler: les sons du langage", les analyses acoustiques apportent des informations limitées quant aux processus de production, particulièrement dans la parole spontanée. Notamment, j'ai fait l'hypothèse que certaines formes de réduction de la parole ne seraient pas strictement sous-spécifiées, mais comporteraient des indices permettant aux auditeurs d'accéder correctement aux informations (socle phonétique). Ces indices sont parfois visibles sur les documents acoustiques sous la forme d'une rupture du processus d'assimilation ou du maintien de détails très fins (Hawkins, 2003). Toutefois, il est probable que des gestes articulatoires soient ébauchés sans trace spécifique sur le signal. Ces gestes pourraient être perceptibles par les auditeurs, malgré tout.

Nous avons la chance, au Laboratoire Parole et Langage, de disposer d'un Articulographe Electro-Magnétique (EMA). J'envisage ainsi d'enregistrer des locuteurs avec ce dispositif au cours d'un entretien (parole spontanée dirigée). Après identification de zones de réduction marquées (particulièrement lors des phénomènes de fusion), il sera possible d'observer les mouvements de la langue au cours de la réalisation de la parole réduite.

Réduction phonétique et fonctions discursives

Nous avons pu mettre en relation la réduction vocalique avec certaines propriétés du lexique (Meunier and Espesser, 2011). Notamment, les mots de fonction montraient une réduction plus forte (sur la durée des voyelles et la répartition spectrale de /a/). Toutefois, la distinction mot de contenu/mot de fonction est assez grossière et ne tient pas compte de l'usage des mots dans une conversation. Ainsi, plus que les propriétés morpho-syntaxique des mots, il me semble pertinent d'observer des propriétés discursives. Par exemple, la conjonction *mais*, considéré comme mot de fonction, est réalisée de façon très hétérogène dans une conversation. Elle peut, tour à tour, et selon sa fonction discursive, être extrêmement réduite ou au contraire très allongée.

En collaboration avec Roxane Bertrand et Laurent Prévot (spécialistes du discours au LPL) nous envisageons donc une analyse phonétique de plusieurs mots de fonction ayant des utilisations variées dans le discours. Il s'agirait de *mais*, *enfin*, *alors*. Ces deux derniers mots sont particulièrement intéressants car, dans leur forme réduite (/fɛ̃/ /lor/), il est fréquent que la première syllabe soit omise, ou très réduite. Cette forme de réduction pourrait être également étudiée dans les investigations articulatoires (EMA, ci-dessus).

V.3 MECANISMES D'ADAPTATION: CONFRONTATION DES SYSTEMES

Le projet TYPALOC, ainsi que l'ensemble des travaux que j'ai menés, ont conforté mais aussi précisé ma conception des mécanismes d'adaptation dans la production de la parole. Mes hypothèses de départ concernant les dysarthries ont été nuancées. Je postulais, au départ, que les locuteurs dysarthriques produisaient une hypoarticulation contrainte tandis que celle des sujets sains était uniquement adaptative (en fonction de la situation de parole). Les résultats semblent montrer que l'hypoarticulation reste adaptative, même lorsque le locuteur est confronté à des contraintes motrices. Un locuteur reste donc toujours dans des stratégies linguistiques (si ces stratégies ne sont pas, elles-mêmes, atteintes).

Ces observations ainsi qu'un échange avec un collègue de Toulouse ont ravivé mon intérêt pour l'étude des productions de locuteurs non-natifs. Au cours d'un échange sur la production phonétique de non-natifs en situation d'apprentissage du français, Michel Billières s'étonnait d'un comportement linguistique de ces apprenants. Il me précisait que, lors du cours, et en situation de production contrôlée de mots isolés, les étudiants produisaient de très bons prototypes du français. Mais dès qu'ils sortaient de cours et qu'ils s'entretenaient librement avec des francophones, leur production redevenait très éloignée d'une production native.

Ce constat renforce l'hypothèse d'un processus d'adaptation optimal concernant le locuteur sain natif. Ce processus serait universel dans sa forme global mais dépendant du système linguistique dans sa réalisation. Mon hypothèse est donc que ces apprenants ne disposent pas du mécanisme d'adaptation articulatoire propre au système du français.

Dans ce cadre j'envisage donc, comme pour le projet TYPALOC, une étude comparative de la production de locuteurs natifs et non natifs (apprenants) dans des types de parole différents. Ces travaux permettraient d'observer le décalage entre les deux populations et de mesurer l'empan de la flexibilité entre les différents types de parole (comme dans Fougeron et al., 2014).

Acquérir les mécanismes de production adaptatifs d'une langue est probablement ce qu'il y a de plus complexe. Parler une langue c'est donc apprendre ses propriétés, mais c'est aussi apprendre ses mécanismes de variation et de réduction.

V.4 PERCEPTION EN PAROLE SPONTANÉE

Je ne suis pas pleinement satisfaite des conclusions généralement tirées des expérimentations comportementales très contrôlées pour les raisons que j'ai évoquées plus haut (essentiellement parce qu'elles ne nous donnent pas des informations complètes sur le traitement de la parole dans une situation naturelle). Tout comme je fais l'hypothèse que la production du locuteur sain natif est caractérisée par un processus d'adaptation, il semble inévitable que le système perceptif soit également conditionné, entre autre, par ce type de processus. Le fait que, dans une situation expérimentale, les réponses des sujets dépendent parfois plus de la tâche et de la situation expérimentale que de l'objet linguistique en est un exemple. Si, dans une expérience, on demande à des sujets de faire une tâche de décision lexicale (mots/non mots) et que les stimuli présentés sont uniquement des mots, les sujets vont probablement en identifier certains comme des mots. Car ils s'attendent à ce qu'on leur présente des deux possibilités. Ils s'adaptent donc à la tâche qui leur est demandée. C'est ainsi que la notion de perception catégorielle a pu être remise en question du fait que la réponse catégorielle des sujets était essentiellement dû au fait que la tâche qui leur était demandée était de nature catégorielle (Massaro and Cohen, 1983). Pour ces raisons, il me semble intéressant de développer des tâches perceptives dont le support est, justement, la parole spontanée, même si cela semble extrêmement délicat étant donné l'absence de contrôle des stimuli.

Comme nous l'avons mentionné précédemment, nous faisons l'hypothèse que, dans la parole, l'information est dispersée dans les différents secteurs linguistiques. Cette dispersion est forte lorsque tous les secteurs linguistiques, voire extra-linguistiques sont présents (conversation) et faible lorsque peu de secteurs linguistiques sont présents (dans la lecture de pseudo-mots, seules les informations phonétique, syllabique et phonotactique sont présentes). Donc, d'une certaine façon, dans le processus de perception de la parole spontanée, il faut *rassembler les pièces du puzzle*. Comment se fait cette opération?

Dans l'optique d'utiliser la parole spontanée comme base de stimuli pour des expérimentations perceptives, j'envisage de reprendre la méthodologie adoptée dans une étude ancienne (Meunier & Floccia, 1997, 1999). Dans un premier temps pour évaluer la stabilité ou robustesse d'un nombre plus important de phonèmes concernant la réduction phonétique. Ensuite, pour

évaluer le degré de contexte (empan de parole présentée aux auditeurs) nécessaire à l'identification des unités infra-lexicales et lexicales en fonction de différents facteurs comme la position dans le mot, la position prosodique ou encore la fonction discursive des mots.

VI. REFERENCES

- Ackerman, H. and W. Ziegler (1991) "Articulatory deficits in Parkinson's dysarthria: an acoustic analysis," *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 54, pp. 1093-1098.
- Adda-Decker, M., Adda, G., and Lamel, L. (2013). Les systèmes de transcription automatique de la parole comme instruments de mesure sur les grands corpus oraux. In *Méthodes et Outils Pour L'analyse Phonétique Des Grands Corpus de Parole*, (Paris: Nguyen, N. & Adda-Decker, M.), pp. 159–202.
- Adda-Decker, M., Delais-Roussarie, E., Fougeron, C., Gendrot, C., and Lamel, L. (2012). La liaison dans la parole spontanée familière : une étude sur grand corpus. *Revue Française de Linguistique Appliquée* XVII-1, 113–128.
- Adda-Decker, M., Gendrot, C., Nguyen, N. (2008). Contributions du traitement automatique de la parole à l'étude des voyelles orales du français. *Traitement Automatique des Langues*, 49 (3), 13-46.
- Audibert, N., Fougeron, C., Fredouille, C., and Meunier, C. (2010). Evaluation d'un alignement automatique sur la parole dysarthrique. In *Actes Des 28èmes Journées d'Etude Sur La Parole*, (Mons, Belgique), pp. 353–356.
- Auzou, P., Rolland-Monnoury, V., Pinto, S., Ozsancak, C. (2007) *Les Dysarthries*, Solal.
- Bartkova, K., and Sorin, C. (1987). A model of segmental duration for speech synthesis in French. *Speech Communication* 6, 245–260.
- Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., and Rauzy, S. (2008). Le CID — Corpus of Interactional Data — Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement Automatique Des Langues* 49, 105–134.
- Best, C. (1995) A Direct Realist View of Cross-Language Speech Perception. In Strange, W. (éd.), *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*. Baltimore, IN, USA : York Press, Inc., 171-204.
- Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 345-360. doi:10.1037/0096-1523.14.3.345
- Bigi, B. (2012). SPPAS: a tool for the phonetic segmentations of Speech. *Proceedings of the International Conference on Language Resources (LREC)* (Istanbul, Turkey), pp. 1748–1755.
- Bigi, B., Meunier, C., Nesterenko, I., Bertrand, R. (2010). Annotation automatique en syllabes d'un dialogue oral spontané, *Actes des Journées d'Etude sur la Parole*, Mons (Belgique), mai 2010, 181-184.
- Bigi, B., Meunier, C., Nesterenko, I., Bertrand, R. (2010). Syllable Boundaries Automatic Detection in Spontaneous Speech, *Proceedings of Language Resource and Evaluation Conference*, Malte, mai 2010, 3285-3292.
- Blache, P. & Meunier, C. (2004). Domaines et propriétés : une description de la répartition de l'information linguistique, *Journées d'Etudes Linguistiques*, Mai 2004, Nantes, p. 197.
- Blache, P. & Meunier, C. (2004). Language as a complex system: the case of phonetic variability, *VI Congreso de Lingüística General*, Mai 2004, Saint-Jacques de Compostelle, Espagne.
- Blache, P., Bertrand, R., Guardioli, M., Guénot, M.-L., Meunier, C., Nesterenko, I., Pallaud, B., Prévot, L., Priego-Valverde, B., Rauzy, S. (2010). A formal annotation model as a preliminary step before annotation scheme: an experiment, *Proceedings of Language Resource and Evaluation Conference*, Malte, mai 2010.
- Blache, P.; Bertrand, R.; Bigi, B.; Bruno, E.; Cela, E.; Espesser, R.; Ferré, G.; Guardioli, M.; Hirst, D.; Magro, E.-P.; Martin, J.-C.; Meunier, C.; Morel, M.-A.; Murisasco, E.; Nesterenko, I.; Nocera P., Pallaud B., Prévot L., Priego-Valverde B., Seinturier J., Tan N., Tellier M., Rauzy S. (2010). Multimodal Annotation of Conversational Data. *Proceedings of Linguistic Annotation Workshop*, 2010 juillet, Uppsala, Sweden. 6 pages.

- Blanche-Benveniste C. (1999) Constitution et exploitation d'un grand corpus , *Revue Française de linguistique appliquée*, vol. IV, n°1, p. 65-74.
- Bradlow, A. R. (2002). Confluent talker- and listener-related forces in clear speech production. In Gussenhoven, C. & Warner, N. (Eds.) *Laboratory Phonology 7*. Berlin & New York: Mouton de Gruyter. Pp. 241-273.
- Browman, C.P., and Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica* 49 (3-4), 155–180.
- Burchfield, L.A., and Bradlow, A.R. (2014). Syllabic reduction in Mandarin and English speech. *The Journal of the Acoustical Society of America* 135, EL270–EL276.
- Chomsky, Noam (1957), *Syntactic Structures*, The Hague/Paris: Mouton
- Content, A., Meunier, C., Kearns, R., Frauenfelder, U.H. (2001). Sequence detection in pseudowords in French: Where is the syllable effect? *Language and Cognitive Processes*, 16, 5/6, 609-636.
- Cosi, P., Falavigna, D., and Omologo, M. (1991). A preliminary statistical evaluation of manual and automatic segmentation discrepancies. In *Proceedings of EuroSpeech 91*, (Genova, Italie), pp. 693–696.
- Côté, M., Morrison, G.S. (2007) The nature of the schwa/zero alternation in French clitics: Experimental and non-experimental evidence, *Journal of French Language Studies*, 17 (2007), pp. 159–186
- Cutler, A., Weber, A. (2007). Listening experience and phonetic-to-lexical mapping in L2, *Proceedings of the 16th International Congress of Phonetic Sciences*, August 2007, Barcelona, Germany, 43-48.
- Darley, F. L & Aronson, A. E. & Brown, J. R. (1975). *Motor Speech Disorders*. Philadelphia: W.B. Saunders.
- Darley, F. L. & Aronson, A.E. & Brown, J.R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, 12: 246-269.
- Dohen, M., Schwartz, J.-L., and Bailly, G. (2010). Speech and face-to-face communication – An introduction. *Speech Communication* 52, 477–480.
- Duez, D. (2003). Acoustic properties of consonant sequences in conversational French speech. *International Congress of Phonetic Sciences*, pp. 2965–2968. Barcelone.
- Duez, D. (2007). Acoustic analysis of occlusive weakening in Parkinsonian French speech. In *Proceedings of International Congress of Phonetic Sciences*, pp. 1–4.
- Dufour, S., & Frauenfelder, U. H. (2010). Phonological neighbourhood effects in French spoken-word recognition. *The Quarterly Journal of Experimental Psychology*, 63(2), 226-238.
- Dupoux, E. (1993). The time course of prelexical processing: The syllabic hypothesis revisited. In *Cognitive Models of Speech Processing*, (Altmann G. and Shillcock, R.), pp. 81–111.
- Durand, J. (2005) Les primitives phonologiques: des traits distinctifs aux éléments., In Nguyen, Noël; Wauquier-Gravelines, Sophie; Durand, Jacques (eds.) *Phonologie et phonétique: Forme et substance*. Traité IC2. (Cognition et Traitement de l'Information). Paris: Hermès, 63-93.
- Durand, J., Laks, B., and Lyche, C. (2003). Le projet “Phonologie du français contemporain” (PFC). *La Tribune Internationale Des Langues Vivantes* 3–9.
- Elman J., McClelland J. (1986) Exploiting lawful variability in the speech wave , *Invariance and Variability in Speech Processes*, Perkell J., Klatt D., eds., LEA, Hillsdale, N.J., p. 360-381.
- Ernestus, M. (2013). The production and comprehension of casual speech., In *Experimental Approaches to Perception and Production of Language Variation (ExAPP 2013)*, (Copenhagen, Denmark),.
- Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142, 27–41.
- Ernestus, M., and Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics* 39, 253–260.
- Ernestus, M., Lahey, M., Verhees, F. & Baayen, R.H. (2006). Lexical frequency and voice assimilation. *The Journal of the Acoustical Society of America* 120, 1040-1051.
- Fant, G. (1973). *Speech, Sounds and Features* (Massachusetts: the MIT Press, Cambridge).
- Farnetani, E. (1997). Coarticulation and connected speech. In *The Handbook of Phonetic Sciences*, (Hardcastle W.J., Laver J., eds.), pp. 371–404.

- Flege, J.E. (1995). Second Language speech learning: Theory findings and problems. In W. Strange (ed.) *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*, 133-272. Timonium MD: York Press.
- Forster, K.I., and Chambers, S.M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior* 12, 627-635.
- Fougeron C., Crevier-Buchman L., Fredouille C., Ghio A., Meunier C., Chevrie-Muller C., Audibert N., Bonastre J.-F., Colazo-Simon A., Delooze C. et al (2010) Developing an acoustic-phonetic Characterization of dysarthric speech in French, in *Proceedings of the 7th International Conference on Language Resources (LREC)*, Malte.
- Fougeron, C. et Audibert, N. (2012) Distorsions de l'espace vocalique : quelles mesures? Application à la dysarthrie. *Actes des Journées d'Etudes sur la Parole*, Grenoble, Juin 2012,
- Fougeron, C., Audibert, N., Fredouille, C., Meunier, C., Gendrot, C., and Panseri, O. (2010). Comparaison d'analyses phonétiques de parole dysarthrique basées sur un alignement manuel et un alignement automatique. In *Actes Des 28èmes Journées d'Etude Sur La Parole*, (Mons, Belgique), pp. 365-368.
- Fougeron, C., Audibert, N., Meunier, C. (2013). La réduction vocalique en parole dysarthrique est-elle sensible au style de parole ?, *5ème Journées de Phonétique Clinique*, Liège, Belgique, octobre 2013.
- Fredouille, C., and Pouchoulin, G. (2011). Automatic detection of abnormal zones in pathological speech. In *Proceedings of the International Congress of Phonetic Sciences*, (Hong Kong, China),.
- Frenck-Mestre, C.; Meunier, C.; Espesser, R.; Holcomb, P.; Daffner, K. (2005). Perceiving Nonnative Vowels: The Effect of Context on Perception as evidenced by Event-Related Brain Potentials., *Journal of Speech Language and Hearing Research*, vol. 48, 1496-1510.
- Frenck-Mestre, C.; PERI, P.; Meunier, C.; Espesser, R. (2010). Perceiving non-native vowel contrasts: ERP evidence of the acquisition of non-native contrasts in late learners. *Proceedings of New Sounds 2010*, 2010 avril-mai, Poznan, POLAND, 119-124.
- Fuchs C. (2002) Place et rôle de la variabilité dans les sciences cognitives , *Invariants et variabilité dans les sciences cognitives*, Lautrey J., Mazoyer B., Van Geert P., eds., Editions de la Maison des Sciences de l'Homme, Paris, p. 157-173.
- Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66, 789-806.
- Galliano, S., Geoffrois, E., Mostefa, D., Choukri, K., Bonastre, J.F., and Gravier, G. (2005). The ESTER Phase II Evaluation Campaign for the Rich Transcription of French Broadcast New. In *9th Interspeech Conference*, (Lisboa, Portugal), pp. 2453-2456.
- Gendrot, C. & Adda-Decker M. (2004). Analyses formantiques de corpus radiophoniques multilingues. *Actes de la conférence MIDL 2004*, Paris.
- Gendrot, C., and Adda-Decker, M. (2005). Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. *Variations* 2, 2-4.
- Gendrot, C., and Adda-Decker, M. (2007). Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages. In *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 1417-1420.
- Ghio, Alain; Pouchoulin, Gilles; Teston, Bernard; Pinto, Serge; Fredouille, Corinne; De Looze, Céline; Robert, Danièle; Viallet, François; Giovanni, Antoine (2012). How to manage sound, physiological and clinical data of 2500 dysphonic and dysarthric speakers? *Speech Communication*, vol. 54. 2012, p. 664-679.
- Goldman, J.Ph., Content, A., Frauenfelder, U.H. (1996). Comparaison des structures syllabiques en français et en anglais. *Actes des XXIèmes Journées d'Etudes sur la Parole*, Avignon.
- Gooskens, C., van Heuven, V.J., van Bezooijen, R., and Pacilly, J.J.A. (2010). Is spoken Danish less intelligible than Swedish? *Speech Communication* 52, 1022-1037.
- Gould, S.J. & Vrba, E. (1982) Exaptation- A missing Term in the Science of Form, *Paleobiology*, Vol.8, No. 1, 4-15.

- Hanique, I., Aalders, E., and Ernestus, M. (2013). How robust are exemplar effects in word comprehension? *The Mental Lexicon* 8, 269–294.
- Hawkins, S. (2003) Roles and representations of systematic fine phonetic detail in speech understanding, *Journal of Phonetics*, 31, 373-405.
- Hazan, V. (2007). Second language acquisition and exemplar theory, *Proceedings of the 16th International Congress of Phonetic Sciences*, August 2007, Saarbrücken, Germany, 39-42.
- Hazan, V., Jeon Kim, J., Chen Y. (2010). Audiovisual perception in adverse conditions: Language, speaker and listener effects, *Speech Communication*, 52, Issues 11–12, 996-1009.
- Jamin, M., Trimaille, C., and Gasquet-Cyrus, M. (2006). De la convergence dans la divergence: le cas des quartiers pluri-ethniques en France. *Journal of French Language Studies* 16, 335–356.
- Johnson, K. (2000) Adaptive dispersion in vowel perception. *Phonetica*, 57, 181-188.
- Johnson, K. (2004). Massive reduction in conversational American English. In Spontaneous Speech: Data and Analysis. *Proceedings of the 1st Session of the 10th International Symposium*, pp. 29–54.
- Joos, M. (1972) The isolation of styles, in J. FISHMAN (éd.) *Readings in the Sociology of Language*, Mouton, La Haye, 185-191.
- Jun, S. & Fougeron, C. (2002) Realizations of accentual phrase in French intonation, *Probus* 14: 147-172.
- Kocjančič Antolík T. & Fougeron C. (2013) Consonant distortions in dysarthria due to Parkinson’s disease, Amyotrophic Lateral Sclerosis and Cerebellar Ataxia. *Interspeech’13*, Lyon, France, 26-29 september 2013.
- Kuhl, P.(1995) Speech Perception. In Minifie, *Introduction To Communication Sciences And Disorders*, 77-148.
- Kuhl, P.K., & Iverson, P. (1995). Linguistic Experience and “the perceptual magnet effect. In W. Strange (ed.) *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*, pp. 121-154. Timonium MD: York Press
- Laaridh I. & Fredouille C. & Meunier C. (accepté). Automatic detection of phoneme-based anomalies in dysarthric speech, Special Issue of *ACM Transactions on Accessible Computing (TACCESS) on Speech and Language Interaction for Daily Assistive Technology (SLPAT)*, (article soumis en avril 2014).
- Labrune, L. (2005). Autour de la syllabe: les constituants prosodiques mineur en phonologie. In *Phonétique et Phonologie: Forme et Substance*, (Paris: Nguyen, N., Wauquier-Gravelines, S., Durand, J.), pp. 95–116.
- Lautrey J., Mazoyer B., Van Geert P., eds., (2002) *Invariants et variabilité dans les sciences cognitives*, Editions de la Maison des Sciences de l’Homme, Paris.
- Lindblom B. (1963) Spectrographic study of vowel reduction , *Journal of the Acoustical Society of America*, 35, p. 1773-1781.
- Lindblom B. (1990) Explaining phonetic variation : a sketch of the hyper- and hypospeech theory , *Speech Production and Speech Modelling*, Hardcastle W.J., Marchal A., eds., Kluwer Academic Publishers, Dordrecht, p. 403-439.
- Lindblom B., Lindgren R. (1985) Speaker-listener interaction and phonetic variation , *PERILUS IV*, Department of Linguistics, University of Stockholm, Suède.
- Lindblom, B. (1986) Phonetic Universals in Vowel Systems. In *Experimental Phonology*. Edited by Ohala J.J. & Jaeger J.J., Academic Press Inc., 13-44.
- Logemann, J., H. Fisher, B. Boshes and E. Blonsky (1978) “Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients,” *Journal of Speech and Hearing Disorder*, vol. 43, pp. 47-57.
- Lubker, J., and Gay, T. (1982). Anticipatory labial coarticulation: experimental, biological and linguistic variables. *Journal of the Acoustical Society of America* 71, 437–448.
- Luce, P. A., Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
- Machac, P., and Skarnitzl, R. (2009). *Principles of Phonetic Segmentation* (Prague: Erudica Ed.).

- Maddieson I., Wright R. (1995) The vowels and consonants of Amis – a preliminary phonetic report , *UCLA Working Papers in Phonetics*, Phonetics Laboratory, Los Angeles, CA, USA, p. 45-65.
- Maddieson, I., Flavier, S., Marsico, E., Coupé, C. & Pellegrino, F., (2013), LAPSyD: Lyon-Albuquerque Phonological Systems Database, *proc. of 14th Interspeech Conference*, Lyon, France, 25-29 August
- Manuel S.Y., Krakow R.A. (1984) Universal and language particular aspects of vowel-to-vowel coarticulation , *Haskins Lab. Stat. Rep. Speech Res. SR-77/78*, p. 69-78.
- Marian, V., Bartolotti, J., Chabal, S., Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE* 7(8): e43230. doi:10.1371/journal.pone.0043230
- Marslen-Wilson, W.D., and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10, 29–63.
- Massaro, D.W., and Cohen, M.M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication* 2, 15–35.
- Mattys, S.L., Davis, M.H., Bradlow, A.R., and Scott, S.K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes* 27, 953–978.
- Mehler J., Dommergues J.-Y., Frauenfelder U.H., Segui J. (1981) The syllable's role in speech segmentation , *Journal of Verbal Learning and Verbal Behavior*, 20, p. 298-305.
- Meunier C. (1997). Voicing assimilation as a cue for cluster identification, *Proceedings of the Fifth European Conference on Speech Communication and Technology*, Rhodes, CDROM, 4 pages.
- Meunier C. (1999). Recovering missing phonetic information from allophonic variations., *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, 4 pages.
- Meunier C., Content A., Frauenfelder U., Kearns R. (1997). The locus of the syllable effect: Prelexical or Lexical?, *Proceedings of the Fifth European Conference on Speech Communication and Technology*, Rhodes, Vol. 5, pp. 2851-2854.
- Meunier C., Floccia C. (1999) Syllabe ou mot : quelle unité permet d'identifier les catégories phonétiques ? , *Actes des 2èmes Journées d'Études Linguistiques Syllabes* , Nantes, p. 87-92.
- Meunier C., Floccia, C. (1997). La spécificité des types de parole pour la perception de la voyelle, *Actes des 1ères Journées d'Études Linguistiques*, La voyelle dans tous ses états, Nantes, 158-163.
- Meunier C., Frenck-Mestre C., Lelekov-Boissard T., Le Besnerais M. (2004). La perception des systèmes vocaliques étrangers: une étude inter-langues, *Journées d'Études sur la parole*, Avril 2004, Fès, Maroc, 377-380.
- Meunier C., Frenck-Mestre C., Lelekov-Boissard T., Le Besnerais M. (2003). Production and perception of foreign vowels: does the density of the system play a role?, *Proceedings of the 15th International Congress of Phonetic Sciences*, August 2003, Barcelona, Spain, 723-726.
- Meunier, C. (2012). Contexte et nature des réalisations phonétiques en parole conversationnelle. In *Proceedings of the Joint Conference JEP-TALN-RECITAL*, Volume 1: JEP, (Grenoble, France: ATALA/AFCP), pp. 1–8.
- Meunier, C. (2013). Phoneme deletion and fusion in conversational speech. In *Experimental Approaches to Perception and Production of Language Variation (ExAPP 2013)*, (Copenhagen, Denmark),.
- Meunier, C., and Espesser, R. (2011a). Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics* 39, 271–278.
- Meunier, C., and Espesser, R. (2011b). Is Phoneme Inventory a Good Predictor for Vocal Tract Use in Casual Speech? *Proceedings of the International Conference on Phonetic Sciences (ICPhS)*, Hong Kong, pp. 1370–1373.
- Meunier, C., and Nguyen, N. (2013). Traitement et analyse du signal de parole. In *Méthodes et outils pour l'analyse phonétique des grands corpus oraux*, (Paris: Lavoisier), pp. 85–119.
- Meunier, C., Legou, T. Vidil, C. & Ortega, M. (2013). Elision des consonnes obstruantes dans la maladie de Parkinson: déficit moteur ou propriété linguistique?, *5ème Journées de Phonétique Clinique*, Liège, Belgique, octobre 2013.

- Meunier, C.; Espesser, R.; Frenck-Mestre, C. (2005). The role of temporal information in vowel perceptual areas: a cross-linguistic study., *Plasticity in Speech Perception*, 2005 juin, Londres, UK.
- Meunier, C.; Espesser, R.; Frenck-Mestre, C. (2006). Aspects phonologique et dynamique de la distinctivité au sein des systèmes vocaliques: une étude inter-langue., *Actes des Journées d'Etude sur la Parole (JEP)*, 2006 juin, Dinard, France, 333-336.
- Meunier, C.; Espesser, R.; Frenck-Mestre, C. (2006). Phonetic variability as a static/dynamic process in speech communication: a cross linguistic study., *10th Conference on Laboratory Phonology*, 2006 juin – juillet, Paris.
- Meunier, C.; Frenck-Mestre, C.; Le Besnerais, M. (2005) Vowel variability in speech production context: a cross linguistic study. *Phonetics And Phonology in Iberia*, Barcelone.
- Meunier, C.; Meynadier, Y.; Espesser, R. (2008). Voyelles brèves en parole conversationnelle. *Actes, Journées d'Etude sur la Parole (JEP)*, 2008 juin, Avignon, France, 97-100.
- Meynadier, Y. (2013). Elements de phonétique acoustique. In *Méthodes et outils pour l'analyse phonétique des grands corpus oraux*, (Paris: Adda-Decker, M. & Nguyen, N.), pp. 25–83.
- Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, 24, 45-64.
- New B., Pallier C., Ferrand L., Matos R. (2001) Une base de données lexicales du français contemporain sur internet: LEXIQUE, *L'Année Psychologique*, 101, 447-462. <http://www.lexique.org>
- Nguyen N. (2001) Rôle de la coarticulation dans la reconnaissance des mots , *L'Année Psychologique*, 101, p. 125-154.
- Nguyen, N., Fagyal, S., and Cole, J. (2004). Perceptual relevance of long-domain phonetic dependencies. In *Actes Des Journées d'Etudes Linguistiques*, (Nantes, France), pp. 173–178.
- Niebuhr O, Meunier C. (2011). The Phonetic Manifestation of French /s#j/ and /j#s/ Sequences in Different Vowel Contexts: On the Occurrence and the Domain of Sibilant Assimilation., *Phonetica*, Volume 68, No 3, 133-160.
- Niebuhr, O., Clayards, M., Meunier, C., Lancia, L. (2011). On place assimilation in sibilant sequences—Comparing French and English, *Journal of Phonetics*, Volume 39, Issue 3, July 2011, Pages 429-451.
- Niebuhr, O., Lancia, L., Meunier, C. (2008). On place assimilation in French sibilant sequences. In *proceedings of the 8th International Seminar on Speech Production (ISSP08)*, December 2008, Strasbourg, France, 221-224.
- Perkell J.S, Klatt D.H., eds. (1986) *Invariance and Variability in Speech Processes*, LEA, Hillsdale, N.J..
- Pinker, S. (1994). *The language instinct* (New York: Morrow).
- Pitt, M.A., Johnson, K., Hume, E., Kiesling, S., and Raymond, W. (2005). The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Communication* 45, 89–95.
- Pluymakers, M. Ernestus, M., & Baayen, H. (2005). Lexical frequency and acoustic reduction in spoken Dutch, *Journal of the Acoustical Society of America*, 118, 2561-2569.
- Pollack, O., Pickett, J.M. (1969) Intelligibility of Excerpts from Fluent Speech: auditory vs structural context, *Journal of Verbal Learning and Verbal Behavior*, 3, 79-84.
- Rossi, M. (1990). Segmentation automatique de la parole: pourquoi? Quels segments? *Traitement Du Signal* 7, 315–326.
- Rouas, J.L., Mayumi Beppu, Martine Adda-Decker (2010) Comparison of Spectral Properties of Read, Prepared and Casual Speech in French, in '2010 *International Conference on Language Resources and Evaluation (LREC 2010)*'.
- Sadat, J., Martin, C. D., Costa, A., & Alario, F. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, 68, 33-58.
- Saussure F. de (1916) *Cours de linguistique générale*, Payot, Paris (réédition, 1976).
- Schouten, M.E.H. and Pols, L.C.W (1979), Vowel Segments in Consonantal Context, *Journal of Phonetics*, 7 (1), 1-23.

- Schwartz J.-L., Boë L.J., Vallée N. (1997) Major trends in vowel system inventories , *Journal of Phonetics*, 25, p. 233-253.
- Sebastián-Gallés, N., Dupoux, E., Seguí, J., and Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language* 31, 18–32.
- Segui J., Frauenfelder U.H., Mehler J. (1981) Phoneme monitoring, syllable monitoring and lexical access , *British Journal of Psychology*, 72, p. 471-477.
- Selkirk, E.O. (1981). English Compounding and the Theory of Word-structure. In *The Scope of Lexical Rules*, (Dordrecht: M. Moortgat, H. Van der Hulst & T. Hoestra (eds.)),.
- Simon, Anne-Catherine ; Auchlin, Antoine ; Avanzi, Mathieu ; Goldman, Jean-Philippe (2009) Les phonostyles: une description prosodique des styles de parole en français. In: Abecassi, M. & G. Ledegen, *Les voix des Français. En parlant, en écrivant*, Peter Lang: Berne, 71-88. 978-3-0343-0170-1.
- Smiljanic, R. and Bradlow, A. R. (2008). Temporal organization of English clear and plain speech. *Journal of the Acoustical Society of America*, 124(5), 3171-3182.
- Stevens, K.N., House, A.S. (1963) Perturbation of Vowel Articulations by Consonantal Context: An Acoustical Study, *Journal of Speech and Hearing Research*, 6 (2), 111-128.
- Torreira, F., Adda-Decker, M., and Ernestus, M. (2010). The Nijmegen Corpus of Casual French. *Speech Communication* 52, 201–212.
- Torreira, F., and Ernestus, M. (2011). Vowel elision in casual French: The case of vowel /e/ in the word *c'était*. *Journal of Phonetics* 39, 50–58.
- Tranel B. (1987) *The Sounds of French. An Introduction*, Cambridge University Press, Cambridge, 232 p.
- Troubetzkoy N.S. (1939) Gründzüge des Phonologie , *Travaux du Cercle Linguistique de Prague*, 7, Traduit par Cantineau J., *Principes de phonologie*, 1970, Klincksieck, Paris.
- Turner, G. & Tjaden, K. (2000) Acoustic differences between content and function words in Amyotrophic Lateral Sclerosis. *JSLHR*, 43 (3), 796-815.
- Vallée N. (1994) *Systèmes vocaliques : De la typologie aux prédictions*, Thèse de Doctorat en Sciences du Langage, Université de Grenoble.
- Vallée, N., Boë, L.-J., and Stefanuto, M. (1999). Typologies phonologiques et tendances universelles. Approche substantialiste. *Linx. Revue des linguistes de l'université Paris X Nanterre* 31–54.
- Van Bergem, D.R. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication* 12, 1–23.
- Vanrullen, T., Blache, P., Portes, C., Rauzy, S., Maeyhieux, J.F., Guénot, M.L., Ba Bellengier, E. (2005) Une plateforme pour l'acquisition, la maintenance et la validation de ressources lexical *Proceedings of the TALN Dourdan*, France, 511-516.
- Warren, R.M., Obusek, C.J. (1971) Speech perception and phonemic restoration, *Perception Psychophysics*, vol.9, 358-362.
- Weismar, G. (1984) "Acoustic description of dysarthric speech: perceptual correlates and physiological inferences," *Seminars in Speech and Language*, vol. 5, pp. 293-313.
- Wesenick, M.B., and Kipp, A. (1996). Estimating the quality of phonetic transcriptions and segmentations of speech signals. In *ICSLP-1996*, pp. 129–132.
- Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics* 38, 329–336.
- Yao, Y. & Meunier, C. (2014). Effects of phonological neighborhood density on phonetic variation: The curious case of French. *14th Laboratory Phonology (LabPhon14)*. July 2014, Tokyo, Japan.
- Yao, Y. & Meunier, C. (2014). Phonetic variation in French conversational speech. *2nd Asia Pacific Corpus Linguistics Conference (APCLC 2014)*. March 2014, Hong Kong, China.