UNIVERSITE PARIS 5 - RENE DESCARTES

INSTITUT DE PSYCHOLOGIE

LABORATOIRE DE PSYCHOLOGIE EXPERIMENTALE, UMR 8581

Olivier Crouzet

Segmentation de la parole en mots et régularités phonotactiques :

Effets phonologiques, probabilistes ou lexicaux ?

Thèse présentée et soutenue publiquement
le 1er Décembre 2000
en vue de l'obtention du grade de Docteur de l'Université Paris 5

Discipline : Psychologie

Sous la direction de Madame le Professeur Nicole Bacri

MEMBRES DU JURY

Pr. Nicole Bacri	Université Paris 5
Pr. Ulrich H. Frauenfelder	Université de Genève
Pr. Daniel Holender	Université Libre de Bruxelles
Dr. Annie Rialland	CNRS, Université Paris 3
Dr. Juan Segui	CNRS, Université Paris 5

UNIVERSITE PARIS 5 - RENE DESCARTES

INSTITUT DE PSYCHOLOGIE

LABORATOIRE DE PSYCHOLOGIE EXPERIMENTALE, UMR 8581

Olivier Crouzet

Segmentation de la parole en mots et régularités phonotactiques :

Effets phonologiques, probabilistes ou lexicaux ?

Thèse présentée et soutenue publiquement
le 1er Décembre 2000
en vue de l'obtention du grade de Docteur de l'Université Paris 5

Discipline : Psychologie

Sous la direction de Madame le Professeur Nicole Bacri

MEMBRES DU JURY

Pr. Nicole Bacri	Université Paris 5
Pr. Ulrich H. Frauenfelder	Université de Genève
Pr. Daniel Holender	Université Libre de Bruxelles
Dr. Annie Rialland	CNRS, Université Paris 3
Dr. Juan Segui	CNRS, Université Paris 5

A Julia et Raphaël, avec tout mon amour…

Cette thèse appartient à Madame le Professeur Nicole Bacri, qui a bien voulu accepter de diriger ce travail. J'ai constamment bénéficié de ses suggestions, de ses conseils, de ses encouragements. J'ai trouvé en elle un directeur toujours disponible et ouvert. Ma dette envers elle est immense. Qu'elle veuille bien trouver ici le témoignage de ma profonde gratitude et de ma respectueuse affection.

J'ai eu la chance de conduire ce travail au Laboratoire de Psychologie Expérimentale de la rue Serpente. La multitude des domaines abordés par ses membres a contribué plus que tout à élargir le champ de mes intérêts. Le laboratoire a changé d'adresse mais mon cœur reste au 28, rue Serpente.

Je veux exprimer ici ma vive reconnaissance à Monsieur Juan Segui, directeur du laboratoire. C'est grâce à son bienveillant appui que je dois d'avoir pu mener à bien cet ouvrage. Le séminaire du laboratoire a constitué un lieu privilégié dans le cadre duquel j'ai eu la chance de profiter des interventions de nombreux orateurs. Les vigoureuses discussions qui y ont été engagées constituent un trésor inestimable. Les réunions du vendredi après-midi ont été un lieu d'échange permanent où toutes sortes de problématiques pouvaient être abordées. Les discussions que j'ai eu la chance d'avoir avec Juan Segui, Pierre Hallé, Ludovic Ferrand et les étudiants de l'équipe 'Psycholinguistique' ont constitué un support permanent à ma réflexion.

Les trois années passées dans le bureau 426 resteront à jamais gravées dans ma mémoire. Je tiens à remercier ceux que j'y ai côtoyés : Alix, Amandine, Anne, Boris, Céline, Dorine, Elsa, Fanny, Frédéric Apoux, Frédéric Isel, Hakima, Juliette, Mathieu, Renaud, Sophie, Véronique, Xavier. Votre présence a fait de mon séjour chez les thésards de la 426 une merveilleuse expérience. Caroline Bey et Sophie Donnadieu ne venaient dans ce bureau que rarement, mais elles l'ont éclairé de leur sourire et de leur gentillesse. Marie-Hélène m'a accueilli au sous-sol du centre Henri Piéron dès mon année de DEA et m'a aidé à construire ma première expérience. Son aide m'a été précieuse.

Dès mon arrivée dans le laboratoire, je me suis senti très proche des membres de l'équipe 'Audition'. Steve McAdams et Carolyn Drake m'ont accordé leur soutien permanent. Ma rencontre avec Christian Lorenzi a été décisive. Je lui dis toute ma gratitude pour les nombreuses marques de bienveillance qu'il m'a témoignées. Dès notre première rencontre, j'ai été enthousiasmé par ses travaux. C'est grâce à lui que je pars pour de nouveaux horizons scientifiques et culturels. Qu'il veuille bien trouver ici la marque de mon profond dévouement.

Mon séjour au Laboratoire de Sciences Cognitives et Psycholinguistique a été essentiel pour l'élaboration ultérieure de mon travail de thèse. Je suis profondément reconnaissant à Monsieur Jacques Mehler de m'avoir accueilli pendant une année. Mes échanges avec Luca Bonnati, Anne Christophe, Stanislas Dehaene, Emmanuel Dupoux, Ghislaine Lambertz et Christophe Pallier ont été particulièrement fructueux. Je garde au fond de ma mémoire le souvenir des moments passés avec Josiane et Thierry.

Monsieur le Professeur Uli Frauenfelder m'a accordé toute sa confiance en m'offrant un poste d'assistant à l'Université de Genève durant ma dernière année de doctorat. Je lui en suis profondément reconnaissant. Les discussions engagées au cours des réunions du laboratoire m'ont été précieuses pour mener à bien cette réflexion. Durant cette dernière année, j'ai côtoyé les étudiants de la Faculté de Psychologie et des Sciences de l'Education de l'Université de Genève : Bruno, Cécile, Dao, Elisabeth, Jean-Philippe, Jeremy, Julie, Roberto, Sébastien, ont fait de mon court séjour à Genève une expérience inoubliable. Je remercie Odile et Fabien qui ont eu la gentillesse de m'héberger avant mon départ et m'ont soutenu dans les derniers moments de la rédaction de cette thèse.

Par leurs enseignements, Monsieur Jean-Didier Bagot, Monsieur le Professeur Claude Bonnet, Madame le Professeur Nicole Bacri et Monsieur Pierre Marquer ont fait naître mon profond intérêt pour l'étude des processus perceptifs et langagiers. Ma dette envers eux est immense.

Jamais je n'aurais pu écrire cette thèse sans la patience et l'amour de mes parents. Ils ont été en permanence à l'écoute de mes désirs, ont accepté que je quitte La Rochelle pour Poitiers, puis Poitiers pour Paris. Ces pages leur sont aussi dédiées.

Gérard et Sylvie m'ont adopté dès le premier jour. Ils ont eu la patience de m'accueillir durant de nombreuses années avant que je ne commence cette thèse. J'ai trouvé en eux une seconde famille et leur en serai éternellement reconnaissant.

Julia et Raphaël m'ont soutenu de leur présence durant ces quatre années. J'ai souvent été absent, en pensées autant que physiquement. Ils ont été à mes côtés au jour le jour, ont vécu mes joies et mes difficultés, mes espoirs et mes renoncements. Leurs sourires et leurs rires me redonnaient confiance. C'est par leur amour que tout ceci est devenu possible.

Je souhaite que cette thèse soit digne de toutes les personnes qui ont rendu ce travail possible par leur présence, leur patience, leurs encouragements, leur amitié et leur amour.

Ce travail a pu être mené à bien grâce au soutien financier de la

Délégation Générale pour l’Armement,
Ministère de la Défense.

« Scholarship is the process by which butterflies are transmuted into caterpillars […] »

Jerry A. Fodor (1983). *The modularity of mind*.

TABLE DES MATIERES

Table des matières 7

Introduction 17

Appariement entre signal acoustique et représentations lexicales 20

1. Du signal acoustique aux représentations linguistiques 20
   - 1.1. Du signal acoustique à l'image auditive 21
   - 1.2. L'appariement entre représentations auditives et phonétiques 25
     - 1.2.1. Variabilité intrinsèque 25
     - 1.2.2. Variabilité extrinsèque 26
   - 1.3. L'appariement entre représentations phonétiques et phonologiques 28
2. Des connaissances qui influencent la perception du signal ? 30

2.1. Le recours à des connaissances lexicales 30
2.1.1. Les données expérimentales 30
2.1.1.1. L'effet du statut lexical sur la catégorisation phonémique 31
2.1.1.2. L'effet du statut lexical sur la détection de phonèmes 32
2.1.1.3. Effets lexicaux sur la restauration phonémique 33
2.1.1.4. Effet lexical dans la compensation de la coarticulation 36
2.1.2. Un modèle autonome des phénomènes d'interaction ? 37
2.1.2.1. Le problème des effets 'lexicaux' dans des non-mots 38
2.1.2.2. Le modèle MERGE 39
2.2. Le recours à des connaissances pré-lexicales 42
2.2.1. Les données expérimentales 42
2.2.1.1. La compensation perceptive de la coarticulation 42
2.2.1.2. Le rôle des contraintes phonotactiques dans l'identification de phonèmes ambigus 43
2.2.1.3. Le rôle des contraintes phonotactiques dans l'identification de phonèmes non-ambigus 45

Segmentation du signal de parole en mots 49

1. La segmentation comme conséquence de l'accès au lexique 50
1.1. Cohort 50
1.1.1. Prédiction des frontières lexicales avant la fin acoustique du mot 51
1.1.2. Point d'Unicité tardif et génération d'une cohorte 53
1.2. Trace 55
1.2.1. Alignement exhaustif des activations lexicales 56
1.2.2. Compétitions entre candidats lexicaux 57
1.2.3. Réalisme des procédures implémentées dans TRACE 58
1.3. Shortlist 60
1.3.1. Représentation de l'ordre par récurrence 60
1.3.2. Réseaux récurrents simples et enchâssement lexical 61
1.3.3. Implémentation de procédures de segmentation prélexicales 62
2. Indices de segmentation prélexicaux 63
2.1. Indices segmentaux 63
2.1.1. En anglais 64
2.1.2. En français 65
2.2. Indices suprasegmentaux 66

2.2.1. Indices métriques 66
2.2.1.1. Accent lexical 67
2.2.1.2. Alternances prosodiques 68
2.2.2. Indices probabilistes 70
2.2.2.1. Les travaux sur les langues artificielles 71
2.2.2.2. Traitement ou mémorisation ? 72
2.2.3. Indices phonotactiques 73
2.2.3.1. Les expériences de détection de syllabe 73
2.2.3.2. Syllabation et détection de phonèmes 75
2.2.3.3. Régularités phonotactiques et *word-spotting* 77
2.2.3.4. Peut-on en déduire un rôle des régularités phonologiques dans les processus de segmentation lexicale ? 79

Syllabe et légalité phonotactique 82

1. Représentation sous-jacente et forme de surface 84
1.1. Représentations sous-jacentes linéaires vs. hiérarchiques 84
1.1.1. L'approche initiale de la phonologie générative 85
1.1.1.1. Représentation interne des phonèmes 85
1.1.1.2. Représentation de la séquence de phonèmes 86
1.1.2. La phonologie autosegmentale 89
1.1.2.1. Représentation interne des phonèmes 90
1.1.2.2. Représentation de la séquence de phonèmes 91
1.2. Processus de choix d'une forme de surface 93
1.2.1. Transformation d'une représentation sous-jacente 93
1.2.2. Sélection de la sortie en fonction des caractéristiques souhaitées 93

2. La syllabe 94
2.1. Description de la notion de syllabe 95
2.1.1. A quoi sert la syllabe ? 95
2.1.2. Qu'est-ce que la syllabe ? 97
2.1.2.1. Les constituants 98
2.1.2.2. Les courbes syllabiques 99
2.1.2.3. La syllabe moraïque 99
2.2. Déterminants de la structuration syllabique 101
2.2.1. Principes d'organisation indépendants des caractéristiques des phonèmes 101

2.2.1.1. Approche par règles 102
2.2.1.1.1. Principe de l'Attaque Obligatoire 102
2.2.1.1.2. Principe de l'Attaque Maximale 102
2.2.1.2. Optimalité des formes syllabiques 103
2.2.2. Principes reposant sur les caractéristiques des phonèmes 104
2.2.2.1. Echelle de sonorité 104
2.2.2.2. Similarité 106
2.2.3. Influences supplémentaires sur la syllabation 108

3. Procédures de syllabation et contraintes phonotactiques 109
3.1. La notion de légalité phonotactique 110
3.1.1. Suites de consonnes : groupes et agrégats 111
3.1.2. Groupe légal vs. groupe tautosyllabique 111
3.2. Qu'est-ce qu'une séquence phonotactiquement légale ? 111
3.2.1. Toute séquence de phonèmes attestée à l'attaque syllabique 112
3.2.1.1. Début de syllabe et début de mot 112
3.2.1.2. Présentation des données 113
3.2.1.3. La notion de déviance 113
3.2.2. Toute séquence de phonèmes fréquente 114
3.3. Syllabation, contraintes phonotactiques et segmentation lexicale 114

Analyse distributionnelle des séquences de consonnes 120

1. Présentation du corpus 122
1.1. Caractéristiques de la base de données et méthode d'analyse 122
1.2. Problèmes posés par le choix de ce corpus 123
1.2.1. Transcription des sons 123
1.2.2. Base de données de mots isolés 125

2. Probabilité d'occurrence indépendante de la position 126
2.1. Méthode d'analyse 127
2.1.1. Fréquence d'occurrence 127
2.1.2. Probabilités pondérées 128
2.1.3. Probabilités transitionnelles 128
2.2. Résultats 130
2.2.1. Moyennes 132
2.2.2. Distributions 138

2.3. Discussion 140

3. Probabilité d'occurrence en début de mot 141
3.1. Méthodes d'analyse 143
3.1.1. Fréquence d'occurrence 144
3.1.2. Probabilités pondérées 144
3.2. Résultats 144
3.2.1. Moyennes 145
3.2.2. Distributions 148
3.3. Discussion 150

4. Réanalyse des données comportementales 151
4.1. Rappel des données 152
4.1.1. Word-spotting 152
4.1.2. Détection de phonèmes 153
4.1.3. Interprétations 154
4.2. Interprétations concurrentes 154
4.2.1. Un phénomène de sélection lexicale sérielle ? 155
4.2.2. Segmentation probabiliste 157
4.2.3. Fréquence et compétitions lexicales 158

Contraintes phonologiques et sélection lexicale 164

1. Expérience 1 : Contraintes phonologiques et sélection lexicale 165
1.1. Méthode 165
1.1.1. Sujets 165
1.1.2. Matériel 165
1.1.3. Procédure 167
1.2. Résultats 168
1.2.1. Analyse globale 169
1.2.1.1. Temps de réaction 169
1.2.1.2. Taux d'erreurs 170
1.2.1.3. Discussion 171
1.2.2. Analyses complémentaires 172
1.2.2.1. Contexte phonologique et fréquence d'usage 172
1.2.2.1.1. Résultats 172
1.2.2.1.2. Discussion 173

1.2.2.2. Contexte phonologique et nombre de compétiteurs 174
1.2.2.2.1. Méthode 175
1.2.2.2.2. Résultats 176
1.3. Discussion 177
2. Expérience 2 : Contrôle en décision lexicale 178
2.1. Méthode 179
2.1.1. Sujets 179
2.1.2. Matériel 179
2.1.3. Procédure 179
2.2. Résultats 180
2.2.1. Temps de réaction 180
2.2.2. Taux d'erreurs 182
2.3. Discussion 182
3. Discussion générale 186

Contraintes phonologiques, régularités probabilistes, compétitions lexicales 192

1. Expérience 3 : Groupes de consonnes de même fréquence 194
1.1. Méthode 195
1.1.1. Sujets 195
1.1.2. Matériel 195
1.1.3. Procédure 197
1.2. Résultats 197
1.2.1. Temps de réaction 198
1.2.2. Taux d'erreurs 199
1.3. Discussion 199
2. Expérience 4 : Estimation du rôle de la fréquence 200
2.1. Méthode 201
2.1.1. Sujets 201
2.1.2. Matériel 201
2.1.3. Procédure 204
2.2. Résultats et discussion 204
2.2.1. Temps de réaction 204
2.2.2. Taux d'erreurs 205

2.2.3. Discussion 205
2.3. Analyses complémentaires 206
2.4. Discussion générale 207

3. Expérience 5 : Augmentation du nombre d'items de remplissage 209
3.1. Méthode 209
3.1.1. Sujets 209
3.1.2. Matériel 209
3.1.3. Procédure 210
3.2. Résultats 210
3.2.1. Temps de réaction 211
3.2.2. Taux d'erreur 211
3.3. Discussion 212

4. Expérience 6 : Structure de listes semi-bloquée par condition 214
4.1. Méthode 215
4.1.1. Sujets 215
4.1.2. Matériel 215
4.1.3. Procédure 215
4.2. Résultats et discussion 216
4.2.1. Temps de réaction 216
4.2.2. Taux d'erreur 217
4.2.3. Analyses complémentaires 217
4.2.3.1. Contexte phonologique et fréquence d'usage 218
4.2.3.2. Contexte phonologique et nombre de compétiteurs 218
4.3. Discussion générale 219

Niveau d'implémentation, Mode d'intégration 223

1. Résumé des résultats obtenus 224
1.1. Effets observés 224
1.1.1. Contraintes phonologiques et sélection lexicale 224
1.1.2. Régularités phonologiques, fréquence et compétitions lexicales 225
1.2. Discussion générale 225
1.2.1. Processus de traitement précoces ou stratégies ? 226
1.2.2. Lien entre régularités phonologiques et propriétés lexicales 227
1.2.3. Mode d'intégration des processus de segmentation 228

2. Niveau d'implémentation 229
2.1. Traitement de l'information ou stratégies post-perceptives ? 229
2.1.1. Conditions d'émergence de l'effet phonologique 229
2.1.2. Caractère écologique des processus impliqués 230
2.1.3. Expérience 7 : Groupes de consonnes de même fréquence 232
2.1.3.1. Méthode 232
2.1.3.1.1. Sujets 232
2.1.3.1.2. Matériel 232
2.1.3.1.3. Procédure 234
2.1.3.2. Résultats 234
2.1.3.2.1. Temps de réaction 234
2.1.3.2.2. Taux d'erreur 235
2.1.3.3. Discussion 235
2.1.4. Expérience 8 : Estimation du rôle de la fréquence 236
2.1.4.1. Méthode 236
2.1.4.1.1. Sujets 236
2.1.4.1.2. Matériel 236
2.1.4.1.3. Procédure 237
2.1.4.2. Résultats 238
2.1.4.2.1. Temps de réaction 238
2.1.4.2.2. Taux d'erreur 238
2.1.5. Discussion générale 239
2.2. Interaction entre variables prélexicales et lexicales 242
2.2.1. Nombre de voisins lexicaux et fréquence des diphones 243
2.2.2. Intégration des informations 245
3. Mode d'intégration 247
3.1. Approches théoriques 248
3.1.1. Influence directe des processus prélexicaux sur les niveaux d'activation lexicale 248
3.1.2. Une proposition différente : focalisation attentionnelle sur des groupements perceptifs dérivés des ruptures dans le signal 252
3.1.2.1. Intégration des informations perceptives 252
3.1.2.1.1. Le rôle des alternances prosodiques en français 252
3.1.2.1.2. L'organisation hiérarchique des pièces musicales 253
3.1.2.1.3. Perspectives dérivées de la théorie de la Gestalt 254
3.1.2.2. Segmentation hiérarchique et niveaux d'activation lexicale 255

3.1.3. Une expérience qui permettrait de tester cette hypothèse 257

Conclusion **259**

Références bibliographiques **261**

Index des illustrations **271**

Index des tableaux **275**

Annexes **I**

Annexe 1 : Liste des groupes de consonnes (et consonnes individuelles) recherchés dans l'analyse distributionnelle présentée au Chapitre 4. Transcription IPA (International Phonetics Association) accompagnée de la transcription propre à BRULEX (Content et al., 1990). I

Annexe 2 : Script en langage Awk pour l'analyse distributionnelle des groupes de consonnes dans la base de données lexicale informatisée BRULEX (Content et al., 1990). III

Annexe 4 : Le même script modifié afin de restreindre l'analyse distributionnelle aux mots pour lesquels une information sur la fréquence d'usage est disponible dans BRULEX (Content et al., 1990). IV

Annexe 4 : Exemple de script permettant de compter le nombre d'occurrences des séquences Consonne-Voyelle dans BRULEX (Content et al., 1990). IV

Annexe 7 : Résultats de l'analyse distributionnelle réalisée sans prendre en compte la position dans les mots. Les données sont triées par catégorie phonétique de groupe puis par probabilité descendante d'occurrence. V

Annexe 9 : Seuils de probabilité du test de Scheffé appliqué sur les données de fréquence d'occurrence et de probabilité transitionnelles obtenues sans prendre en compte la position dans les mots. XII

Annexe 11 : Script permettant de compter le nombre d'occurrences de chaque groupe de consonnes en début de mot. XIII

Annexe 13 : Le même script modifié afin de restreindre l'analyse distributionnelle aux mots pour lesquels une information sur la fréquence d'usage est disponible dans BRULEX (Content et al., 1990). XIV

Annexe 15 : Exemple de script permettant de compter le nombre d'occurrences des séquences Consonne-Voyelle apparaissant en début de mot dans BRULEX (Content et al., 1990). XIV
Annexe 17 : Résultats de l'analyse distributionnelle conduite sur les débuts de mots. Les données sont triées par catégorie phonétique de groupe puis par probabilité descendante d'occurrence. XV
Annexe 11 : Seuils de probabilité du test de Scheffé appliqué sur les données de fréquence d'occurrence en début de mot. XXII
Annexe 20 : Consigne lue aux sujets pour les tâches de *word-spotting*. XXIII
Annexe 21 : Matériel utilisé dans l'Expérience 1. XXIII
Annexe 22 : Matériel utilisé dans l'Expérience 3. XXVI
Annexe 23 : Matériel utilisé dans l'Expérience 4. XXVII
Annexe 16 : Matériel utilisé dans l'Expérience 5. XXVIII
Annexe 17 : Matériel utilisé dans l'Expérience 6. XXIX
Annexe 26 : Matériel utilisé dans l'Expérience 7. XXIX
Annexe 27 : Matériel utilisé dans l'Expérience 8. XXX
Annexe 20 : Items de remplissage utilisés dans l'expérience 1. XXX
Annexe 21 : Items de remplissage utilisés dans les expériences 3 et 6. XXXI
Annexe 22 : Items de remplissage utilisés dans l'expérience 4. XXXI
Annexe 23 : Items de remplissage utilisés dans l'expérience 5. XXXI
Annexe 24 : Items de remplissage (*fillers*, à gauche) et 'attrapes' (*foils*, à droite) utilisés dans l'expérience 7. XXXII
Annexe 25 : Items de remplissage (*fillers*, à gauche) et 'attrapes' (*foils*, à droite) utilisés dans l'expérience 8. XXXIII

INTRODUCTION

Tout être humain ayant accès au langage parlé intègre, au cours de son développement, un ensemble de 'règles' déterminant l'organisation des sons dans sa langue. Cette connaissance implicite des contraintes d'organisation des sons permet au locuteur de transformer une *représentation phonologique sous-jacente* -représentation abstraite de la séquence de phonèmes à prononcer- en un signal de parole présentant des caractéristiques propres à sa langue -*la forme phonétique de surface*. Ainsi, la représentation phonologique du mot français /medəsɛ̃/ sera prononcée [mɛtsɛ̃] en raison des contraintes d'assimilation du /d/ en [t] et d'ouverture du /e/ dans une syllabe fermée. Cet ensemble de règles, désigné sous le terme de 'grammaire' de la langue, constitue une 'base de connaissances' qui pourrait s'avérer d'un intérêt essentiel dans le déroulement des processus d'identification et de segmentation d'un signal de parole. Le type de contraintes auxquelles nous nous sommes intéressé dans ce travail de thèse porte sur les séquences de phonèmes 'admissibles'. Dans chaque langue, certaines séquences phonémiques sont prohibées à l'intérieur d'une syllabe. Ces contraintes, dites phonotactiques, déterminent en partie la structuration syllabique du signal de parole. Dans cette optique, nous nous sommes attaché à étudier le rôle éventuel de ces contraintes, qui sont un des indices disponibles pour la segmentation syllabique, dans les processus de segmentation du signal de parole en mots.

Nombre de travaux effectués en psychologie ont conduit à affirmer que les locuteurs d'une langue donnée peuvent avoir recours à ce type de connaissances pour identifier (Chapitre 1) et / ou segmenter (Chapitre 2) un signal de parole. L'objet de notre travail est d'appréhender ces travaux à travers une vision critique de la validité des données avancées comme reflétant le recours, de la part des locuteurs, à des connaissances sur ce qu'est une forme 'phonologiquement légale' dans leur langue. Après avoir présenté d'une part les travaux qui semblent mettre en évidence le rôle de connaissances dans les processus de traitement de la parole et, d'autre part, les diverses interprétations que l'on peut proposer, nous nous attacherons à décrire le lien entre légalité phonotactique et syllabation (Chapitre 3). Du fait des problèmes posés par les différentes définitions de la légalité, nous serons conduits à choisir une définition opérationnelle à partir d'une analyse distributionnelle de l'occurrence des groupes de consonnes dans un lexique français (Chapitre 4). Nous nous attacherons alors à analyser les données avancées comme preuve de l'importance de ces contraintes dans les processus de segmentation de la parole en mots. Cette analyse nous permettra de montrer que plusieurs modèles d'accès au lexique pourraient prédire les résultats obtenus sans avoir recours à une segmentation prélexicale fondée sur des contraintes phonologiques séquentielles. Par la suite, un certain nombre d'expériences seront présentées dans lesquelles nous avons contrôlé le matériel utilisé afin de dissocier les diverses interprétations possibles des effets obtenus (Chapitres 5 et 6). Enfin, nous nous poserons la question du niveau d'implémentation de ces contraintes dans un modèle de perception de la parole et des relations entre les différents niveaux de traitement (Chapitre 7).

Chapitre 1

Appariement entre signal acoustique et représentations lexicales

APPARIEMENT ENTRE SIGNAL ACOUSTIQUE ET REPRESENTATIONS LEXICALES

1. Du signal acoustique aux représentations linguistiques

La parole est un signal physique produit par la mise en mouvement d'un nombre considérable d'organes (poumons, cordes vocales, langue, lèvres, etc.). Les modifications rapides de la configuration des articulateurs dans le tractus vocal donnent lieu à des frottements qui, selon le modèle source-filtre ('source-filter model' ; Fant, 1960; Flanagan, 1972), produisent une onde. La répercussion de cette onde sur les diverses parois (buccale, nasale, …) du conduit vocal provoque l'amplification de certaines parties du spectre (ce qui équivaut à l'application d'un filtre). On appelle *formants* les composantes fréquentielles de la voix qui subissent une amplification. Certains phonèmes peuvent correspondre à une amplification de la quasi-totalité du spectre perceptible par l'humain ou à une évolution rapide de ces composantes, c'est le cas de la plupart des consonnes. Les voyelles quant à elles sont caractérisées par une structure spectrale relativement stable dans le temps. Le signal acoustique résultant de ces phénomènes mécaniques se caractérise par une organisation spectro-temporelle complexe qui, par l'intermédiaire du milieu de transmission (en général aérien), est transmise au système auditif périphérique par des phénomènes de transduction mécanico-électriques et transformée en un *percept auditif* (Delgutte, 1987). Afin d'aboutir à un *percept linguistique*, le système perceptif doit apparier cette image auditive avec des représentations linguistiques abstraites phonémiques ou phonologiques, forme sous laquelle seraient représentés les mots dans le lexique mental.

1.1. Du signal acoustique à l'image auditive

Du fait de sa complexité, plusieurs étapes fonctionnelles de traitement sont nécessaires avant d'aboutir à une représentation linguistique d'un signal de parole. La première étape, réalisée par les organes de l'oreille interne, consiste à analyser ce signal acoustique (transformé en un signal électrique par les récepteurs sensoriels de l'oreille externe) afin d'en extraire une information qui pourra être utilisée par le système nerveux central. L'un des organes les plus importants pour le traitement de la parole et l'accès à des représentations linguistiques est la cochlée, qui est supportée par l'organe de Corti. Apte à décoder les différentes fréquences d'un son en temps réel (par l'intermédiaire des cellules ciliées internes) et dotée d'une sélectivité fréquentielle considérable (en raison de l'action des cellules ciliées externes), elle constitue un banc de filtres qui fournit au système auditif une analyse fréquentielle *et* temporelle des signaux acoustiques sensiblement comparable à ce que l'on peut observer sur un spectrogramme (Moore, 1997). On considère classiquement que l'information primordiale pour la perception de la parole est constituée par la sortie de ce filtre, qui fournit en temps réel une information sur l'évolution

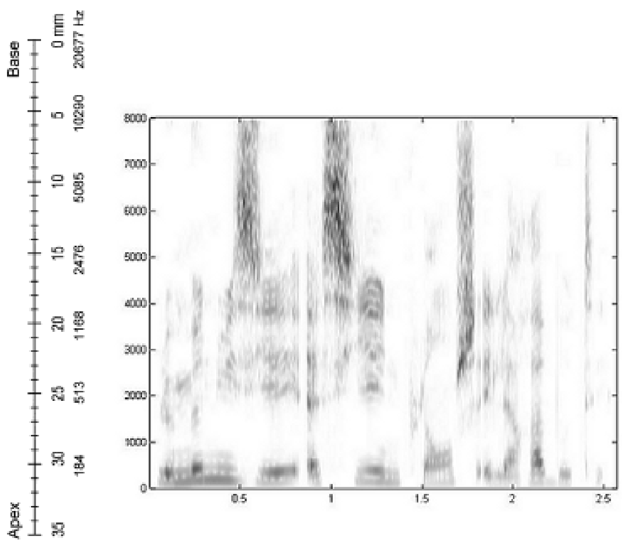

Figure 1 : Illustration du codage tonotopique effectué par la cochlée schématiquement représentée sur la gauche du graphique. Le traitement effectué peut en partie s'assimiler à une analyse en temps réel de l'enveloppe spectrale, le signal de sortie correspondant alors approximativement à ce que l'on observe sur un spectrogramme.

des différentes composantes fréquentielles du signal en fonction du temps (cf. Figure 1). Chaque phonème de la langue peut être décrit par des caractéristiques spectrales spécifiques (Stevens, 1998). L'une des tâches du système d'identification de la parole consiste donc à apparier ces patterns spectraux à des représentations phonétiques. Selon Klatt (Lexical Access From Spectra, 1979 ; 1989), le système de traitement de la parole comparerait chaque spectre à court-terme avec des patterns spectraux prototypiques et rechercherait celui qui correspond le mieux au calcul effectué sur le signal. Chaque phonème de la langue serait lié à un nombre considérable de patterns prototypiques. L'appariement entre le produit des calculs et les représentations spectrales prototypiques stockées en mémoire permettrait alors d'avoir accès à des représentations phonétiques. Stevens (1960 ; 1996) accorde une importance essentielle au lien entre perception et production pour rendre compte du traitement de la parole mais se démarque des propositions avancées dans le cadre des théories motrice (Liberman & Mattingly, 1985) et directe-réaliste (Fowler, 1986). Pour dériver une représentation phonétique du signal acoustique, il propose de relier le système perceptif à un synthétiseur vocal. Ce synthétiseur générerait des spectres à court terme à partir d'un ensemble de règles de production du signal de parole. Ces règles de production permettraient de relier une représentation phonétique à un ensemble de paramètres articulatoires. Le système comparerait le spectre calculé avec chacun des spectres produits par le synthétiseur. Il évaluerait alors la distance entre le spectre calculé et le spectre produit afin de modifier les paramètres du synthétiseur. Cette étape serait répliquée jusqu'à ce que le synthétiseur produise un spectre suffisamment proche du spectre perçu. Il serait alors très facile d'identifier le phonème correspondant au spectre traité puisque le spectre généré par le synthétiseur correspondrait à un programme articulatoire (donc phonétique) connu.

Mais cette représentation spectrale du signal acoustique ne constitue pas la seule information utile à ces processus d'appariement entre signal acoustique et représentations linguistiques (Van Tasell, Soli, Kirby, & Widin, 1987 ; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Le rôle de l'enveloppe d'intensité (des modulations temporelles à long terme) dans l'identification des phonèmes a ainsi pu être mis en évidence grâce à des méthodes de traitement du signal qui consistent à présenter à des auditeurs une bande de bruit lissée avec la forme de l'enveloppe d'intensité d'un signal de parole (cf. Figure 2). Dans cette situation il reste possible, malgré l'absence de toute information spectrale fine, d'accéder à une représentation linguistique du signal. Par exemple, après un entraînement intense, on peut observer des taux d'identification correcte de phrases qui avoisinent les 50 % (Shannon et al., 1995). Dans une tâche de choix forcé à 16 alternatives dans laquelle les auditeurs doivent identifier la consonne médiane de logatomes[1] VCV[2], les auditeurs sont en mesure d'atteindre des taux d'identification correcte de l'ordre de 20 % (Apoux, Berthommier, Bacri, & Lorenzi, 1998) alors que le taux de réponses au hasard correspond à 6,25 %[3]. Que l'on considère la situation de compréhension de phrases ou d'identification de logatomes, ces stimuli ne fournissent au système auditif aucune information spectrale puisqu'ils ne sont en fait que des séquences de bruit blanc dont l'intensité évolue au cours du temps. Mais les modulations d'intensité sonore de ce bruit permettent d'accéder au moins partiellement à une représentation phonétique du signal. Cette aptitude des auditeurs humains à utiliser les indices fournis par les modulations temporelles à long terme présentes dans l'enveloppe constitue la preuve que la forme de l'enveloppe d'intensité d'un signal acoustique apporte des informations importantes pour l'identification de la parole -en tout cas pour la détection de certains indices linguistiques-, donc pour l'appariement entre signal acoustique et représentations linguistiques.

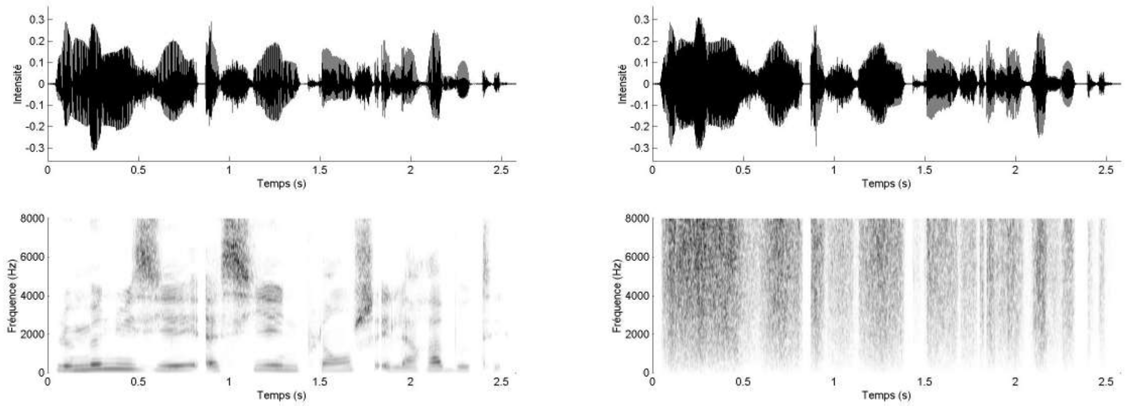

Figure 2 : La phrase 'Le menuisier a scié une planche et l'a rabotée' ; à gauche dans sa forme originale et à droite après suppression des composantes spectrales fines.

Il a également été mis en évidence qu'il n'est pas nécessaire, pour aboutir à un percept structuré, de disposer d'une information précise quant au moment d'occurrence des différents événements fréquentiels du signal de parole dans le temps. Si l'on applique une désynchronisation artificielle au signal de parole (cf. Figure 3), on observe une aptitude des auditeurs à conserver une compréhension correcte des messages linguistiques malgré des taux de réverbération considérables entre bandes de fréquence (Greenberg & Arai, 1998). Cette

---

1 Un logatome est une séquence de parole courte et sans signification, comme par exemple /aka/.

2 VCV : Séquence de parole présentant une structure Voyelle - Consonne - Voyelle.

3 La variabilité des taux de performance en fonction de la tâche est certainement déterminée par le type d'entraînement auquel sont soumis les participants. Dans les expériences de Shannon et al. (1995), les participants avaient déjà entendu les phrases dans leur forme originale. La quantité de phonèmes prononcés doit certainement rendre plus faciles les processus d'appariement entre le signal perçu et la représentation linguistique stockée en mémoire. Plus que les indices temporels disponibles dans l'enveloppe globale, il est possible d'envisager qu'en

désynchronisation constitue l'une des manifestations des phénomènes de réverbération qui sont en fait bien réels dans les signaux auxquels nous sommes confrontés dans la plupart des situations naturelles de communication[4] et consiste à introduire décalages de phase entre différentes bandes de fréquence du signal de parole. Greenberg & Arai (1998) montrent que ces ruptures de synchronie ne gênent pas considérablement l'identification de phrases pour un décalage de phase inférieur à 150 ms. Ce phénomène met en évidence l'existence d'une fenêtre d'intégration temporelle qui permettrait au système auditif de structurer un signal acoustique malgré les distorsions qui peuvent être appliquées au signal acoustique dans de nombreuses situations. Cette capacité d'intégration dans l'accès à des représentations linguistiques montre qu'il n'est pas nécessaire de dériver du signal acoustique une représentation spectrographique précise et que certains événements acoustiques peuvent être décalés dans le temps lorsqu'ils parviennent au système auditif périphérique sans pour autant déstabiliser les percepts qui en découlent.

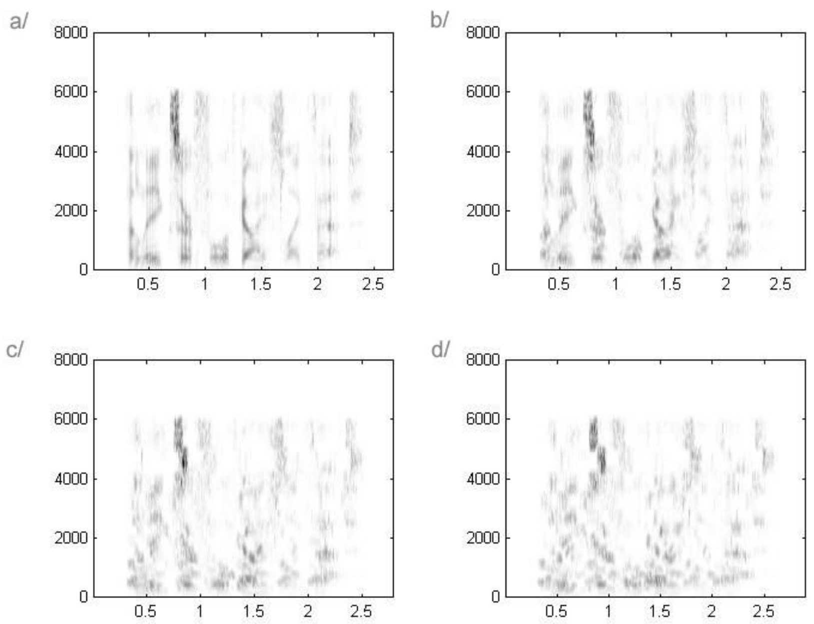

Figure 3 : Désynchronisation artificielle de la phrase 'The leagues are full of guys like that' (d'après Greenberg & Arai, 1998). Chaque spectrogramme correspond à une durée moyenne de désynchronisation de a/ 0 ms (signal original), b/ 40 ms, c/ 120 ms, d/ 220 ms. Les stimuli correspondants sont accessibles sur le site `http://www.ICSI.berkeley.edu/~steveng`

réalité les modulations d'amplitude de l'enveloppe sont traitées à l'intérieur de chaque bande de fréquence (cf. le rôle des cartes de Modulation d'Amplitude ou *Amplitude Modulation maps*, Greenberg & Araï, 1998).

[4] Ceci s'explique par le fait que, dans un environnement ouvert, chaque fréquence du spectre est renvoyée avec une vélocité différente par les divers objets qui sont présents (en raison de leur taille, de leur forme et des caractéristiques du matériau qui les constitue).

On peut donc affirmer que plusieurs types d'images auditives (évolution des composantes fréquentielles en fonction du temps, forme de l'enveloppe d'intensité, intégration temporelle sur des segments d'image auditive de type spectrographique) peuvent être pris en compte par les processus de traitement auditif pour l'accès à des représentations linguistiques.

1.2. L'appariement entre représentations auditives et phonétiques

Dès lors que l'une ou plusieurs de ces images auditives ont été générées par le système auditif périphérique, le système cognitif va devoir apparier chaque portion temporelle de ces images avec des représentations linguistiques abstraites (matrices de traits, phonèmes, diphones, syllabes). Or les signaux de parole que doit traiter le système auditif humain présentent une variabilité importante (Blumstein, 1986 ; Klatt, 1986). Il est possible de définir deux types de variabilité. Nous appelons variabilité intrinsèque les formes de variabilité qui affectent le signal de parole sans être déterminées par le contexte phonétique (hauteur de la voix, vitesse d'élocution). La variabilité extrinsèque correspond aux formes de variabilité qui sont déterminées par le contexte. Les problèmes posés par ces diverses sources de variabilité pourraient éventuellement trouver leur solution dans des procédures communes (par exemple l'analyse en ondelettes, cf. infra.).

1.2.1. Variabilité intrinsèque

L'une des principales difficultés posées par cette étape est liée à l'importante variabilité des productions possibles pour un même message linguistique. Cette variabilité est à la fois fréquentielle et temporelle. Du fait des différences dans la taille des cavités articulatoires des locuteurs (liées à leur âge, à leur sexe, etc.), leurs fréquences de résonance diffèrent d'un individu à l'autre. Les filtres qui vont amplifier ou atténuer certaines fréquences du signal source sont donc différents. Chaque locuteur génère ainsi des signaux de parole avec une hauteur de voix (fréquence fondamentale ou F_0) différente ; ce phénomène induit également une répartition variable des formants sur l'échelle des fréquences. Par ailleurs, un locuteur peut parler plus ou moins rapidement en fonction des situations et changer de vitesse d'élocution à l'intérieur d'un même énoncé. Cette variabilité dans l'organisation spectrale et temporelle des sons de parole n'entrave cependant pas la stabilité perceptive qui permet à un auditeur d'entendre des phonèmes formes stables.

L'une des méthodes qui ont été proposées afin de rendre compte de la capacité du système cognitif humain à gérer la variabilité du signal de parole consiste à effectuer une normalisation spectrale et / ou temporelle. Cette méthode repose sur le principe de l'appariement entre un signal physique de forme variable et un référent. Par des méthodes mathématiques qui consistent à transformer la représentation spectrographique du signal à la fois dans le domaine spectral et temporel, on peut apparier des signaux de parole variant sur ces deux dimensions avec une représentation normalisée. Ce type de procédure pose cependant le problème des méthodes adéquates qui permettent d'*identifier* le signal référent pertinent. On trouve dans le domaine des travaux sur la vision des propositions alternatives telles les *cônes généralisés* ('generalized cones', Marr, 1982 p.223) ou les *géons* ('geometric ions', Biederman, 1987). Le principe consiste à utiliser un nombre restreint de formes (un *cône* chez Marr, diverses formes de base nommées *géons* chez Biederman) pour décrire ou définir une scène visuelle. Si l'on prend l'exemple des *cônes généralisés*, une forme conique peut être utilisée pour reproduire intégralement une scène visuelle par la combinaison d'un nombre considérable de ces unités. En faisant varier leur taille et leur disposition respectives, on peut générer une reproduction de la scène visuelle réelle. Il reste alors à identifier en mémoire le ou les objets qui correspondent à cette combinaison de cônes. Le problème de la normalisation -donc de l'identification adéquate de l'objet sur la base duquel on effectuera cette normalisation- ne se pose pas puisque la taille des cônes contribuant à l'image n'a d'importance que relative : c'est la configuration des cônes entre eux qui permet de récupérer l'objet en mémoire. Dans le domaine du traitement de la parole, des propositions similaires tendent à voir le jour actuellement avec l'utilisation des transformations en ondelettes ('wavelets transformation', Graps, 1995) comme une alternative à la Transformée de Fourier qui est utilisée actuellement pour aboutir à une représentation spectrographique, mais le recours à ces procédures se limite encore aux travaux effectués en Reconnaissance Automatique de la Parole, et rien n'a été proposé jusqu'à maintenant, comme l'avaient fait Marr (1982) ou Biederman (1987) pour la vision, dans le cadre de la description d'un modèle du fonctionnement cognitif appliqué au traitement de la parole.

1.2.2. Variabilité extrinsèque

Les difficultés posées par l'appariement entre image(s) auditive(s) et représentations abstraites proviennent également des contraintes inhérentes à la coproduction de segments proches. En effet, si l'on admet que les représentations linguistiques auxquelles devra être apparié le signal acoustique sont des phonèmes -donc des unités segmentales discrètes-, ce signal n'est pas, du fait de ce que l'on appelle la *coarticulation*, constitué d'une séquence discrète de segments. Ainsi, chaque unité identifiée par un auditeur est extraite d'une portion de signal dont les caractéristiques dépendent aussi de la prononciation des segments qui l'environnent (cf. Figure 4). Ce phénomène de coarticulation est lié au mode de production de la parole qui consiste à préparer les mouvements articulatoires correspondant à un segment alors même que l'on est en train d'articuler le segment qui précède (effets d'anticipation). Les effets coarticulatoires se manifestent aussi par persévérance ; ainsi, nous générons des mouvements destinés à produire un son alors que l'on n'a pas encore terminé de produire le son précédent. Ceci pose deux énigmes essentielles qui sont intimement liées l'une à l'autre : la première est celle de la segmentation du signal en unités dont la taille correspond à un segment phonémique. Dans une séquence CV, ce phénomène de coarticulation induit la présence simultanée d'informations correspondant à la consonne et d'autres correspondant à la voyelle. Comment découper le signal en segments représentant chacun une unité puisque ce signal n'est pas discret ? La seconde est induite par le fait que ce caractère coarticulatoire des sons de parole génère une extrême variabilité dans la réalisation des sons en fonction de leur voisinage. Les formes acoustiques correspondant au phonème /g/ ne sont pas les mêmes selon qu'il est suivi d'un /i/ ou d'un /a/. Comment, dès lors, proposer des procédures d'appariement entre image auditive et représentation linguistique puisque les manifestations acoustiques des phonèmes sont si variables ?

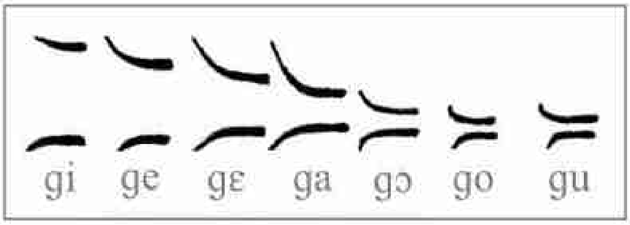

Figure 4 : Schéma représentant le phénomène de coarticulation. On peut voir l'influence des voyelles sur la prononciation du phonème /g/ (d'après Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

Outre les procédures d'appariement spectral et temporel d'un signal donné à une forme auditive référente, les études portant sur la compréhension de la parole doivent donc expliquer comment l'image auditive -une fois normalisée ou décrite à partir d'indices non-sensibles aux variations spectro-temporelles- peut être appariée à des représentations phonétiques. Pour cela, on a cherché à décrire un certain nombre de phénomènes stables dans le signal afin de pouvoir prédire, à partir d'un événement acoustique, le segment phonétique qui lui correspond. Dans cette optique, de nombreux travaux ont été effectués à partir des années cinquante dans lesquels on a tenté de mettre au jour des indices sur lesquels pourrait se fonder le système cognitif afin d'identifier avec certitude les différents phonèmes de la chaîne parlée. Ces indices constitueraient des invariants de relations acoustico-phonétiques qui permettraient de définir, avec une plus ou moins grande certitude, la relation entre forme acoustique et représentation linguistique. Les travaux réalisés dans ce domaine ont eu pour objectif de décrire des caractéristiques invariables permettant de faire le lien entre une forme acoustique et un trait distinctif (voisement, mode d'articulation par exemple). Ils ont conduit à proposer plusieurs indices pour chaque type d'opposition phonémique. Par exemple, le trait de voisement est lié au délai qui sépare l'occlusion consonantique du début de vibration des cordes vocales mais aussi aux caractéristiques de l'enveloppe spectrale de la consonne (Summerfield & Haggard, 1977). Selon Stevens & Blumstein (1978), c'est la contribution conjuguée de la forme des transitions formantiques et du spectre à court terme de la consonne qui permet l'identification du mode d'articulation. Pour certaines valeurs de l'un des indices, il est nécessaire de disposer de l'autre indice pour identifier le mode d'articulation. Cette variabilité du signal de parole s'accompagne par ailleurs d'une redondance considérable des indices acoustiques (Stevens, Keyser, & Kawasaki, 1986). Il est donc possible, même si l'un des indices met le système en échec, d'aboutir à une identification correcte du phonème. La redondance des indices permet en effet d'appliquer des procédures probabilistes de prise de décision qui peuvent tolérer une certaine déviance par rapport à la forme phonétique idéale.

1.3. L'appariement entre représentations phonétiques et phonologiques

Indépendamment des difficultés générées par la variabilité du signal de parole en termes de distributions spectrale et temporelle des événements acoustiques pour la segmentation et l'identification phonétiques, les contraintes inhérentes aux langues posent des difficultés supplémentaires. Chaque langue est soumise à un ensemble de contraintes qui déterminent la réalisation effective des sons de parole (phénomènes d'allophonie, d'assimilation, contraintes phonotactiques portant sur les séquences de phonèmes admissibles). Supposons par exemple qu'un locuteur souhaite prononcer la phrase :

/ʒœprɑ̃ləmetroasizœrpuraleʃeləmedəsɛ̃/

« Je prends le métro à six heures pour aller chez le médecin »

Si l'on admet que les mots sont représentés dans le lexique sous une forme phonologique abstraite correspondant à ce que les phonologues appellent la *représentation sous-jacente*, ce message a peu de chances d'être prononcé de manière aussi canonique. Au contraire, la phonologie et la phonétique montrent que de nombreuses contraintes vont transformer la forme canonique de ce message en une 'suite de segments phonétiques' qui, dans le cadre d'un modèle de compréhension de la parole, ne sont pas directement appariables avec les représentations phonologiques sous-jacentes. La plupart des locuteurs, en fonction bien sûr des conditions de production (lecture vs. parole spontanée), mais aussi du style de langue (soutenue ou pas), vont aboutir à une forme phonétique qui différera considérablement de la représentation phonologique de départ. Cette séquence pourrait par exemple donner lieu à la suite phonétique suivante :

[ʃprɑ̃lmetroasizørpuraleʃelmɛtsɛ̃]

On notera la séquence initiale [ʃp] qui vient se substituer à la représentation phonologique /ʒœp/ par l'élision du /œ/ et l'assimilation de /ʒ/ à [ʃ]. Or il est couramment admis, malgré l'existence de positions alternatives qui fournissent des arguments convaincants[5] (cf. par exemple Goldinger, 1998), que l'unité de contact avec les représentations lexicales devrait être, dans un souci d'économie de traitement, la plus abstraite possible. En effet, si l'on parvient à représenter les mots stockés dans le lexique avec un nombre minimal d'unités de codage, on réalise une économie considérable par rapport à un stockage qui se ferait sous forme acoustique, auquel cas l'ensemble des exemplaires acoustiques de chaque mot devrait être stocké dans le lexique afin de pouvoir le reconnaître. Ainsi, si l'on admet que les mots doivent être représentés dans le lexique sous la forme la plus abstraite qui soit (c'est à dire la représentation phonologique sous-jacente proposée par la phonologie), l'existence de contraintes déterminant les modifications à apporter à la forme phonologique pour aboutir à une réalisation phonétique considérablement différente constitue un obstacle supplémentaire aux processus de compréhension de la parole, non seulement pour l'appariement entre représentations phonétiques et phonémiques (phénomènes d'allophonie nécessitant l'intervention de procédures spécifiques afin d'apparier les différentes variantes phonétiques d'un même phonème, par exemple les différents /r/ du français : [ʁ] et [r]) mais également pour la mise en correspondance des représentations phonémiques et phonologiques (assimilation du trait de voisement par exemple, modification de la qualité des voyelles en fonction de la structure syllabique, contraintes phonotactiques, …). Au-delà des problèmes posés au système perceptif pour l'appariement entre image auditive et représentation phonétique, les contraintes phonologiques introduisent donc des difficultés supplémentaires pour la compréhension de la parole. L'une des tâches du système de perception de la parole va donc consister, une fois le décodage acoustico-phonétique effectué, à apparier cette représentation phonétique avec une représentation phonologique adéquate afin d'être en mesure de contacter les représentations lexicales stockées en mémoire.

[5] Et qui démontrent qu'il est probablement nécessaire d'adopter une position intermédiaire entre un modèle dans lequel le stockage se ferait intégralement sous forme abstraite et un autre dans lequel l'ensemble des exemplaires acoustiques possibles serait représenté.

2. Des connaissances qui influencent la perception du signal ?

Comme nous venons de le voir, l'accès à une représentation linguistique du signal de parole s'avère particulièrement complexe et passe par la mise en œuvre de nombreuses étapes fonctionnelles de traitement. Afin de rendre compte de la possibilité d'appariement entre forme auditive et représentations phonétiques ou phonologiques, de nombreux auteurs envisagent l'utilisation, à l'intérieur du système de reconnaissance de la parole, de connaissances de haut niveau qui rétroagiraient sur des niveaux de représentation moins élaborés (représentation segmentale du signal de parole notamment). Ces connaissances porteraient sur les caractéristiques qui constituent la langue et qui se développent au cours de l'acquisition : le lexique et les contraintes phonologiques.

2.1. Le recours à des connaissances lexicales

Afin de pouvoir identifier correctement les phonèmes dans le signal de parole, l'une des solutions qui ont été envisagées pour faciliter l'appariement d'une image auditive avec des représentations phonémiques ou phonologiques consiste à faire intervenir des procédures de rétroaction des niveaux lexicaux vers les niveaux de représentation prélexicaux. Ainsi, des connaissances de haut niveau faciliteraient la tâche du système d'identification phonémique en guidant les choix perceptifs dans les situations problématiques. Les partisans d'une approche interactive (notamment McClelland & Elman, 1986) se sont heurtés à un courant autonome dont les représentants (Cutler, Mehler, Norris, & Segui, 1987) affirmaient l'indépendance entre niveaux de traitement de bas niveau et représentations plus élaborées. Nous décrivons ici certaines des données expérimentales qui ont été présentées comme reflétant des preuves du recours à des rétroactions lexicales. Une rapide description des interprétations alternatives qui ont été proposées est présentée pour chacun des effets. Ceci nous conduira à mettre au jour l'un des problèmes essentiels dans l'étude du rôle de certaines caractéristiques des langues dans les processus perceptifs : la confusion entre diverses variables qu'il peut s'avérer difficile de contrôler du fait même des particularités de la langue. La conscience de cette confusion sera essentielle pour notre travail sur le rôle des contraintes phonologiques dans les processus de segmentation lexicale.

2.1.1. Les données expérimentales

Quatre types d'effets seront présentés ici qui ont donné lieu à une discussion sur l'existence de processus rétroagissant des représentations lexicales vers les représentations phonémiques : l’effet du statut lexical (1) sur la catégorisation de phonèmes ambigus, (2) sur les temps de détection de phonèmes, (3) sur la restauration phonémique (Warren, 1970) et (4) sur la compensation perceptive de la coarticulation (Elman & McClelland, 1988).

2.1.1.1. L’effet du statut lexical sur la catégorisation phonémique

L’un des effets interprété comme une preuve de l’interaction lexique-phonèmes est celui, mis en évidence par Ganong (1980), du statut lexical sur la catégorisation de phonèmes ambigus. Cet auteur a présenté à des auditeurs des séquences de parole CVC constituées d’une consonne occlusive initiale qui variait sur un continuum de voisement (temps d’attaque vocale -en anglais VOT (Voice Onset Time)- qui distingue /d/ de /t/ par exemple). Ces phonèmes ambigus étaient présentés dans des séquences de type ‘$^{d}_{t}$VC’[6] qui pouvaient donner lieu à un mot pour une extrémité du continuum (/t/ précédant /æsk/ donne *task*, ‘tâche’) et à un non-mot pour l’autre extrémité (/d/ précédant /æsk/ donne le non-mot *dask*). Le contexte variait afin de contrebalancer la consonne qui donnait lieu à l’extrémité lexicale. Ainsi, dans la condition opposée, le phonème ambigu était présenté dans le contexte /æʃ/ qui donnait lieu à un mot après le phonème /d/ (*dash*, ‘tiret’) et à un non-mot après /t/ (/tæʃ/). Ganong met en évidence un déplacement de la frontière catégorielle qui s’exprime par une préférence, de la part des participants, pour des réponses en accord avec une interprétation lexicale de la séquence de phonèmes dans les parties médianes du continuum. Dans le contexte /æsk/, les auditeurs ont plutôt tendance à classer le phonème ambigu dans la catégorie /t/ alors que l’attitude inverse se manifeste dans le contexte /æʃ/ qui donne lieu à une proportion plus élevée de réponses /d/. Cet effet met en évidence une tendance à percevoir un signal de parole en faveur d’une interprétation lexicale plutôt que non lexicale.

Une interprétation *autonome* peut cependant en être donnée. Dans le modèle Race (Cutler & Norris, 1979), les décisions phonémiques peuvent se faire à partir de deux classes d’informations elles-mêmes dérivées de deux voies de traitement : une voie prélexicale et une voie lexicale. On peut simuler des effets lexicaux par la combinaison et l’intégration de ces deux classes d’informations lors de l’étape de ‘prise de décision’. Ainsi, dans le cas de l’effet observé par Ganong (1980), la voie lexicale peut tout à fait influencer l’étape décisionnelle en

[6] Classiquement, dans ce type d’expériences, le phonème ambigu est noté /?/. Nous avons choisi de ne pas suivre ce ‘standard’ afin, d’une part de ne pas causer de confusion avec le caractère /ʔ/ (occlusive glottale non-voisée) de l’International Phonetics Association (IPA) mais aussi pour permettre au lecteur de garder à l’esprit les 2 éléments extrèmes du continuum considéré. Ainsi, le phonème ambigu correspondant à un continuum acoustique qui va du phonème /d/ au phonème /t/ est noté /$^{d}_{t}$/.

introduisant un biais à donner une réponse phonémique conforme aux informations dérivées du traitement lexical. La tâche des sujets consistant à classifier un son dans une catégorie phonémique, la voie non-lexicale fournit une information purement acoustique sur laquelle peut se fonder le sujet pour apparier ce son avec une représentation phonétique et effectuer une catégorisation[7]. Parallèlement, les informations fournies par le signal permettent d'engendrer un traitement lexical. Au moment de la prise de décision, le sujet dispose de deux types d'informations pour donner sa réponse : une information acoustico-phonétique qui peut ne pas suffire à effectuer la catégorisation pour les items centraux du continuum et une information lexicale qui favorise nécessairement -d'autant plus que l'information acoustique n'est pas 'fiable'- l'une des deux possibilités étant donné qu'une seule des deux extrémités correspond à un mot de la langue. Dans ce modèle, c'est cette information qui va influencer la réponse du sujet et induire un effet lexical ; mais à aucun moment du traitement proprement dit, il n'y a action des niveaux de représentations lexicaux sur les niveaux de représentation phonémiques. L'interprétation des résultats obtenus avec la tâche d'identification phonémique a donné lieu à une quantité importante de travaux qui ne permettent pas de trancher entre ces deux interprétations (cf. Pitt & Samuel, 1993 pour une revue).

2.1.1.2. L'effet du statut lexical sur la détection de phonèmes

Une preuve du recours aux connaissances lexicales dans les processus d'identification phonémique a également été entrevue dans certains effets que l'on a pu mettre en évidence avec la tâche de détection de phonèmes. Dans ces expériences, les auditeurs devaient détecter un phonème qui pouvait apparaître dans diverses positions (au début, au milieu, à la fin) de stimuli auditifs lexicaux ou non. L'idée essentielle est que si les niveaux de traitement lexicaux rétroagissent vers les étapes sublexicales de représentation, on devrait observer des effets différents en fonction du statut lexical et du taux d'avancement de l'identification lexicale. C'est effectivement ce qui a été observé dans plusieurs expériences. On a notamment pu mettre en évidence un effet de la position du phonème à détecter qui diffère selon le statut lexical des items porteurs plurisyllabiques. Dans des mots plurisyllabiques, on n'observe pas d'effet lexical sur les temps de détection de phonèmes lorsque le phonème-cible est en position initiale (Foss & Blank,

[7] On pourra objecter qu'il est nécessaire de disposer d'informations sublexicales afin de rendre possible l'accès lexical. On considèrera en fait l'information linguistique déterminant l'accès à cette représentation lexicale comme une information différente de celle dérivée des traitements réalisés par la voie sublexicale (par exemple, des matrices de traits ou de spectres à court terme pourraient générer l'accès lexical sans donner lieu à une identification phonémique proprement dite, …). On peut trouver dans Marslen-Wilson & Warren (1994) un exemple de modèle utilisant des matrices de traits non-discrètes pour l'appariement avec le lexique. Les mêmes informations sublexicales pourraient servir à développer des représentations phonémiques sans que celles-ci aient pour autant une influence sur l'identification des mots.

1980 ; Segui, Frauenfelder, & Mehler, 1981). Par contre, un effet lexical émerge lorsque le phonème-cible apparaît en fin de plurisyllabe (Cutler, Mehler et al., 1987 ; Frauenfelder, Segui, & Dijkstra, 1990) ; les phonèmes sont alors détectés plus rapidement dans des mots que dans des non-mots.

Ces effets pourraient également être prédits par un modèle purement autonome de la tâche de détection de phonèmes dans lequel l'information lexicale permettrait de fournir des informations supplémentaires pour la prise de décision et rendre par conséquent plus rapide la génération de la réponse du sujet. On observe cependant que ces effets lexicaux augmentent progressivement au fur et à mesure que le phonème-cible est localisé vers la fin du mot. On n'observe pas, contrairement à ce que l'on aurait pu prédire à partir d'un modèle dans lequel aucun processus de propagation de l'activation lexicale n'est implémenté, une absence totale d'effet avant le *Point d'Unicité* (PU) suivie d'une brusque émergence après le PU (Frauenfelder et al., 1990). Il y a au contraire une évolution progressive de ces processus rétroactifs qui est corrélative des phénomènes d'activation lexicale et qui n'est pas prédite par un modèle purement symbolique tel que Race (Cutler & Norris, 1979) dans lequel l'information lexicale n'est disponible qu'un fois le mot identifié.

2.1.1.3. Effets lexicaux sur la restauration phonémique

Des effets lexicaux émergent également dans la tâche de restauration phonémique. Le phénomène de restauration phonémique a été mis en évidence par Warren (1970). La tâche consiste à présenter à des sujets deux types de stimuli : des séquences de parole dans lesquelles l'un des segments a été remplacé par du bruit et d'autres dans lesquelles on a ajouté du bruit au segment, celui-ci restant indemne. Si l'on demande aux auditeurs de décider pour chaque stimulus si le segment bruité a été remplacé ou s'il est indemne, ils éprouvent du mal à donner une réponse correcte. Cette difficulté induit les auditeurs à percevoir un segment de parole qui a été remplacé par du bruit comme relativement indemne. Les sujets réagissent comme si le bruit ne faisait que recouvrir le signal de parole alors que celui-ci est en réalité absent. Ce phénomène repose en partie sur des processus ascendants de traitement de l'information que l'on peut observer dans les traitements psychoacoustiques (cf. la continuité illusoire ; Bregman, Colantonio, & Ahad, 1999). Un effet lexical a cependant été mis en évidence par Samuel (1996) qui se réfère à la Théorie de la Détection du Signal (TDS ; Green & Swets, 1966). Si l'on demande à un sujet de réaliser la tâche décrite précédemment, on peut observer un phénomène de restauration phonémique qui consiste à distinguer difficilement les stimuli intacts des stimuli dans lesquels le signal a été remplacé par du bruit. Ceci se manifeste par une tendance à percevoir le segment remplacé par du bruit comme intact. Cet effet peut se mesurer avec une variable dérivée de la TDS : le *d'*. Cette variable est considérée comme un indice de la discriminabilité perceptive entre conditions expérimentales. Samuel (1996) met en évidence un phénomène de restauration phonémique pour les deux types de stimuli (mots ou non-mots) ; mais ce phénomène s'accentue pour les mots : la valeur du *d'* est moins élevée pour les mots que pour les non-mots. En d'autres termes, les sujets éprouvent plus de difficultés à distinguer les stimuli intacts des stimuli dans lesquels le signal a été supprimé lorsqu'ils identifient des mots que lorsqu'ils traitent des stimuli qui n'ont pas de représentation spécifique dans le lexique. Cette diminution, dans les mots par rapport aux non-mots, de la discriminabilité perceptive entre un segment auquel on a *ajouté* du bruit et un segment qui a été *remplacé* par du bruit a été interprétée par Samuel (1996) comme un indice d'une rétroaction du lexique vers les représentations phonémiques. Là encore, on peut objecter que ce phénomène ne reflète pas nécessairement une rétroaction lexico-phonémique. Le problème se situe ici au niveau de la signification -et surtout de l'interprétation- du *d'*. Dans le cadre de la TDS, le *d'* est un indice de discriminabilité perceptive. Un effet en termes de *d'* représente une différence dans les processus perceptifs mais pas dans des processus stratégiques (qui se refléteraient plutôt dans la variable β). Si l'on se réfère à un modèle dans lequel les effets lexicaux sont simulés par une intégration des informations sublexicales et lexicales, les informations lexicales peuvent contribuer à influencer la prise de décision, mais ceci ne coïncide pas nécessairement avec le développement d'une stratégie de réponse[8]. Les effets observés sur la variable *d'* peuvent donc tout à fait refléter un processus d'intégration d'informations et se refléter dans les valeurs de la variable *d'*. On ne peut par conséquent pas inférer de ces modifications du *d'* que ces effets lexicaux sont liés à une rétroaction lexique-phonème qui induirait une modification du codage même de l'information phonémique, plutôt qu'à une étape d'intégration des deux classes d'informations qui seraient codées indépendamment l'une de l'autre.

Plus récemment, Samuel (1997) a mis en évidence la possibilité, dans une procédure d'adaptation sélective, de générer un effet lexical dans le phénomène de restauration phonémique. La procédure d'adaptation consiste à soumettre des auditeurs à l'écoute prolongée de séquences de parole dans lesquelles l'un des phonèmes est prononcé de manière répétée (par

[8] On utilisera donc ici le terme 'intégratif' pour désigner des processus d'intégration des diverses informations disponibles à la sortie du traitement -intégration qui peut se faire par pondération des différentes catégories d'information ou par sélection de l'information jugée la plus pertinente- ; et on les distinguera des processus 'stratégiques' ou 'décisionnels' qui, dans le cadre de la théorie de la détection du signal, correspondraient à l'élaboration (consciente ou non) d'une *ligne de conduite* dans le choix des réponses afin d'accomplir au mieux la tâche avec le moins d'erreurs possibles (ce qui correspondrait alors au critère β).

exemple [p]). Au bout d'un certain temps, on fait écouter aux auditeurs des stimuli qui contiennent un phonème ambigu situé sur un continuum acoustique (imaginons que ce continuum corresponde à une modification du temps d'attaque vocale et que, par conséquent, le phonème opposé à celui qui était présenté dans la phase d'adaptation soit [b]). Dans une seconde phase, on présente des stimuli ambigus qui varient sur un continuum [$^{b}_{p}$]. Le phonème [$^{b}_{p}$] peut être interprété soit comme celui qui s'est répété au cours de la phase d'adaptation ([p]), soit comme celui qui correspond à l'autre extrémité du continuum ([b]). On observe une tendance, suite à la phase d'adaptation, à interpréter le phonème ambigu comme faisant partie de la classe opposée à celle à laquelle les auditeurs ont été soumis au préalable (Eimas & Corbit, 1973). Une saturation des récepteurs qui interviendrait à différents niveaux de traitement (sensoriel, phonétique, phonémique ; Samuel & Kat, 1996) se mettrait en place qui induirait un déplacement de la frontière perceptive. Dans la situation imaginée par Samuel (1997), les sujets sont soumis dans une première phase à une procédure d'adaptation phonémique. Ils entendent, au cours de cette période, des stimuli dans lesquels c'est toujours le même phonème qui est détérioré (par remplacement ou superposition). Ces stimuli peuvent correspondre à des mots ou à des non-mots. L'idée avancée par Samuel (1997) est que, si les représentations lexicales rétroagissent effectivement sur les niveaux de représentation prélexicaux en modifiant l'activation de ces derniers, l'effet lexical qui accentue le phénomène de restauration phonémique devrait induire un autre phénomène : une adaptation phonémique sélective à l'une des extrémités du continuum, en l'occurrence celle qui donne lieu à un mot de la langue. On devrait alors observer, dans la seconde phase de l'expérience, une tendance à percevoir le phonème ambigu dans la catégorie opposée à celle du phonème qui avait été remplacé par du bruit dans la première phase. Effectivement, les résultats mettent en évidence une tendance, après une période d'adaptation au cours de laquelle les auditeurs entendaient des mots dans lesquels l'un des phonèmes était remplacé par du bruit, à favoriser le phonème opposé sur le continuum. Selon Samuel (1997), ceci met en évidence l'action effective du lexique sur les représentations phonémiques prélexicales par une modification de l'activation des récepteurs phonémiques. On peut cependant, là encore, affirmer qu'un modèle autonome peut rendre compte de cet effet. La question se pose notamment de savoir à quel niveau se situent précisément le phénomène d'adaptation phonémique et l'influence lexicale observés par Samuel (1997). On peut se demander si les représentations phonémiques sur lesquelles sont censées agir les informations lexicales sont réellement des représentations précoces recevant des activations de niveaux de représentation plus élaborés ou si, au contraire, ces représentations phonémiques ne sont pas calculées parallèlement aux informations lexicales, auquel cas une interprétation similaire à celle qui peut être proposée pour les effets observés dans la tâche de catégorisation phonémique est envisageable. Si représentations lexicales et sublexicales sont calculées par le biais de deux voies de traitement différentes, le phénomène d'activation des unités phonémiques par les unités lexicales que l'on observe dans le phénomène de restauration phonémique peut refléter un processus post-lexical lié à une intégration des deux types d'informations. L'adaptation induite par la restauration phonémique constituerait alors, elle aussi, un processus décisionnel dans lequel les récepteurs phonémiques auraient tendance à voir leur seuil d'activation s'élever en raison de cette influence qu'à le lexique sur les phonèmes lors de l'étape d'intégration des informations.

2.1.1.4. Effet lexical dans la compensation de la coarticulation

Finalement, l'effet essentiel sur lequel repose l'affirmation d'une boucle rétroactive du lexique vers les représentations phonémiques est celui mis en évidence par Elman & McClelland (1988). Cet effet repose sur un processus nécessairement pré-lexical et mis en évidence par (Mann & Repp, 1981) : la compensation perceptive des phénomènes coarticulatoires (*compensation-for-coarticulation*). En anglais, un phonème ambigu [$^{t}_{k}$] a tendance à être perçu [t] s'il est prononcé après [ʃ] alors qu'il est plutôt perçu [k] s'il est prononcé après [s]. Ce phénomène de compensation perceptive trouve son explication dans le partage de la place d'articulation des occlusives [t] et [k] et, respectivement, des fricatives [s] et [ʃ], partage qui induirait à compenser cette communauté de place d'articulation par une tendance à percevoir le phonème suivant dans la catégorie opposée. Les auditeurs utiliseraient donc une connaissance qu'ils auraient intégrée sur les phénomènes coarticulatoires afin d'influencer le processus d'appariement entre image auditive et représentation phonétique. Elman & McClelland (1988) utilisent ce phénomène dans des séquences lexicales qui se terminent par un phonème ambigu provenant d'un continuum [s]-[ʃ], ce phonème étant lui-même suivi d'un autre phonème présentant une ambiguïté entre [t] et [k]. Les sujets pouvaient par exemple entendre la séquence [fulɪ$^{s}_{\text{ʃ}}$] suivie d'un mot commençant par [$^{t}_{k}$]. Si le mot *foolish* provoque une modification de l'activation du phonème /ʃ/ par un processus de rétroaction du lexique vers une représentation prélexicale, on devrait alors mettre en évidence un phénomène de compensation perceptive induisant les auditeurs à percevoir de manière privilégiée le phonème ambigu [$^{t}_{k}$] comme un représentant de la catégorie /t/. Au contraire, si le biais qui était observé dans les expériences de catégorisation phonémique (Ganong, 1980) est localisé au niveau de l'extraction d'une information phonémique à partir d'une représentation lexicale, la voie prélexicale de traitement devrait fournir des informations non-biaisées du phonème ambigu [^{s}ʃ]. On ne devrait alors observer aucune préférence pour l'un des deux phonèmes /t/ ou /k/ en ce qui concerne le second phonème ambigu. Or les auteurs observent au contraire un déplacement de la frontière catégorielle du son [^{t}k] en faveur d'une compensation perceptive de la coarticulation. Lorsque les auditeurs traitent la séquence /fulɪ^{s}ʃ^{t}k/ composée d'une suite de deux phonèmes ambigus, l'interprétation lexicale favorise un percept /ʃ/ pour le premier phonème, ceci influence alors la perception du second phonème dans la catégorie /t/. Les informations lexicales sont donc en mesure de favoriser l'augmentation d'activation du phonème /ʃ/ qui, à son tour, modifie l'activation du phonème /t/. Il existe cependant, à nouveau, une possibilité d'interprétation alternative. En anglais, en effet, il apparaît que la séquence /ɪʃ/ est plus fréquente que la séquence /ɪs/. On peut alors envisager que l'influence exercée par le contexte [ɪ] sur la catégorisation du phonème ambigu [^{s}ʃ] est purement prélexicale et liée à des phénomènes probabilistes (cf. Saffran, Newport, & Aslin, 1996; Vitevitch & Luce, 1999 ; ces travaux seront présentés plus en détails dans le Chapitre 3) ; auquel cas l'effet qui est observé dans cette expérience pourrait s'expliquer par des processus qui sont tous localisés à un niveau pré-lexical. On n'aurait alors aucune nécessité de postuler une rétroaction lexique-phonèmes dans les processus de perception de la parole : deux processus prélexicaux se succéderaient, le premier consistant à favoriser l'interprétation du [^{s}ʃ] dans la catégorie [ʃ] du fait même de la plus grande probabilité de [ɪʃ] en anglais ; le second -conséquence du biais probabiliste à percevoir [ʃ]- consistant à faire appel à des 'connaissances articulatoires' pour favoriser la perception d'un [t] dans le continuum [^{t}k].

2.1.2. Un modèle autonome des phénomènes d'interaction ?

Concernant l'existence de processus rétroactifs permettant aux représentations lexicales d'influencer le codage même des représentations prélexicales, la question reste en suspens. Dans les quatre exemples qui précèdent, nous avons vu qu'un modèle autonome ne pouvait être exclu pour rendre compte des données considérées jusqu'ici comme des preuves de l'interaction entre niveaux de traitement plus ou moins élaborés. Il reste cependant des résultats que ne peut prédire le modèle RACE (Cutler & Norris, 1979).

2.1.2.1. Le problème des effets 'lexicaux' dans des non-mots

Un certain nombre d'auteurs ont mis en évidence l'existence d'effets 'lexicaux' dans des expériences dans lesquelles les stimuli étaient tous des non-mots. Connine, Titone, Deelman, & Blasko (1997), utilisant une tâche de détection de phonème dans des non-mots, observent des temps de réaction plus rapides dans les non-mots qui ressemblent à des mots ('gabinet' vs. 'cabinet' par exemple) que dans des non-mots qui ne ressemblent à aucun mot. Newman, Sawusch, & Luce (1997), quant à eux, ont répliqué l'expérience de Ganong (1980) en n'utilisant que des non-mots aux deux extrémités des continua. Les paires de non-mots différaient sur le plan du nombre de voisins lexicaux (i.e. de la densité de voisinage). L'une des extrémités constituait un non-mot qui avait beaucoup de voisins lexicaux (par exemple /gaɪs/ ou /kaɪp/) alors que l'autre avait peu de voisins lexicaux (par exemple /kaɪs/ ou /gaɪp/). Ils observent un phénomène similaire à celui observé par Ganong (1980) : il y a déplacement de la frontière catégorielle vers la catégorie phonémique qui donne lieu à une faible densité de voisinage. En d'autres termes, les auditeurs ont tendance à donner une réponse qui favorise le phonème correspondant au non-mot qui a la plus forte densité de voisinage.

Ces effets mettent en évidence l'existence de processus d'accès au lexique (ou plus précisément de propagation de l'activation lexicale) au cours du traitement de stimuli de parole qui n'ont pas de représentation dans le lexique. Ces processus influencent l'interprétation qui peut être donnée, par les participants, de stimuli phonétiques ambigus. Un modèle comme Race (Cutler & Norris, 1979) ne peut pas prédire ce type d'effets. Dans ce modèle, les non-mots sont uniquement traités par le biais de la voie prélexicale. Comme ils n'ont pas de représentations dans le lexique et que Race n'implémente pas de processus de propagation de l'activation lexicale qui pourrait être induite par des portions de non-mots qui existent dans des mots de la langue, cette voie lexicale ne peut pas influencer -même au cours de l'étape décisionnelle- les réponses des sujets lorsqu'ils traitent des non-mots. Un modèle comme TRACE (McClelland & Elman, 1986) pourrait au contraire prédire cet effet en raison de la mise en œuvre de procédures de propagation progressive de l'activation. Il n'est cependant pas exclu que les effets liés à la densité lexicale des non-mots (Newman et al., 1997) soient interprétables en termes pré-lexicaux plutôt que rétroactifs. En effet, il existe une corrélation forte entre le nombre de voisins lexicaux d'une chaîne de phonèmes et la fréquence des diphones qui la constituent (Vitevitch, Luce, Pisoni, & Auer, 1999; Vitevitch & Luce, 1999).

2.1.2.2. Le modèle MERGE

Norris, McQueen, & Cutler (2000) proposent un modèle qui permet de rendre compte des effets de compétition lexicale observés dans des non-mots tout en n'implémentant aucune boucle de rétroaction du lexique vers les phonèmes. Ce modèle est un Réseau Récurrent Simple (Simple Recurrent Network, SRN ; Jordan, 1986) qui, contrairement à TRACE (McClelland & Elman, 1986) dans lequel les connexions sont définies *a priori* par l'expérimentateur (*Interactive Activation and Competition model, IAC*) subit une phase d'apprentissage au cours de laquelle il régule lui-même le poids des différentes connexions entre les nœuds du réseau. On peut alors parler de *Réseau Neuronal* à proprement parler en raison des capacités du modèle à apprendre : c'est l'organisation des poids des différentes connexions qui détermine le comportement du modèle à l'issue de la phase d'apprentissage. Les réseaux récurrents simples (SRN), du fait de l'implémentation de procédures de récurrence locales -des unités cachées (*Hidden Units*) vers les unités d'entrée- autorisent l'insertion d'un retard temporel dans le retour de l'information et permettent ainsi d'aboutir à une représentation correcte des aspects séquentiels (et temporels) des informations traitées (Jordan, 1986 ; Elman, 1990) mais peuvent aussi, grâce à l'étape d'apprentissage, développer des représentations sensibles aux régularités séquentielles des formes rencontrées (cf. Chapitre 2 pour une présentation de cet aspect). Ces modèles n'utilisent que des connexions ascendantes[9] ; aucune rétroaction d'un niveau de représentation élaboré vers un niveau moins élaboré n'est possible. Le modèle MERGE est en fait le *cousin* connexionniste de Race (Cutler & Norris, 1979). Dans le même esprit, deux types d'informations[10] sont disponibles à la sortie du modèle (lexicales et sublexicales) ; et la même étape d'intégration de ces deux classes d'informations est implémentée. Ce qui diffère est en fait la possibilité qu'a le modèle de propager progressivement l'information lexicale vers les nœuds phonémiques -qui rappelons-le sont à considérer ici comme reflétant des représentations de nature différente mais de niveau similaire, les deux processus se déroulant en parallèle- sans attendre l'aboutissement du processus d'accès lexical.

Norris et al. (2000) montrent que ce modèle est en mesure de simuler toutes les données qui étaient présentées comme des preuves de l'existence d'une rétroaction du lexique sur les phonèmes. Il peut en outre, et contrairement à Race (Cutler & Norris, 1979), reproduire les effets de voisinage lexical observés dans des non-mots. Si ce modèle peut reproduire des effets d'interaction, c'est que les informations lexicale et sub-lexicale sont représentées comme deux

[9] La boucle de récurrence ne constitue pas une boucle de rétroaction proprement dite puisqu'elle a pour fonction de donner une mémoire temporelle au modèle et pas de moduler le codage effectué à un autre niveau du traitement.
[10] On parle de *Multiple Output models* (Boland & Cutler, 1996).

classes équivalentes d'unités de sortie. Ce sont donc les procédures décisionnelles -ou d'intégration des informations- (comme dans Race, Cutler & Norris, 1979) qui permettent l'influence de l'une sur l'autre. Du fait de sa nature connexionniste, il implémente également des processus de propagation progressive de l'activation des diverses unités. Il est ainsi en mesure de reproduire tous les effets dont nous avons parlé précédemment sans introduire de boucle rétroactive.

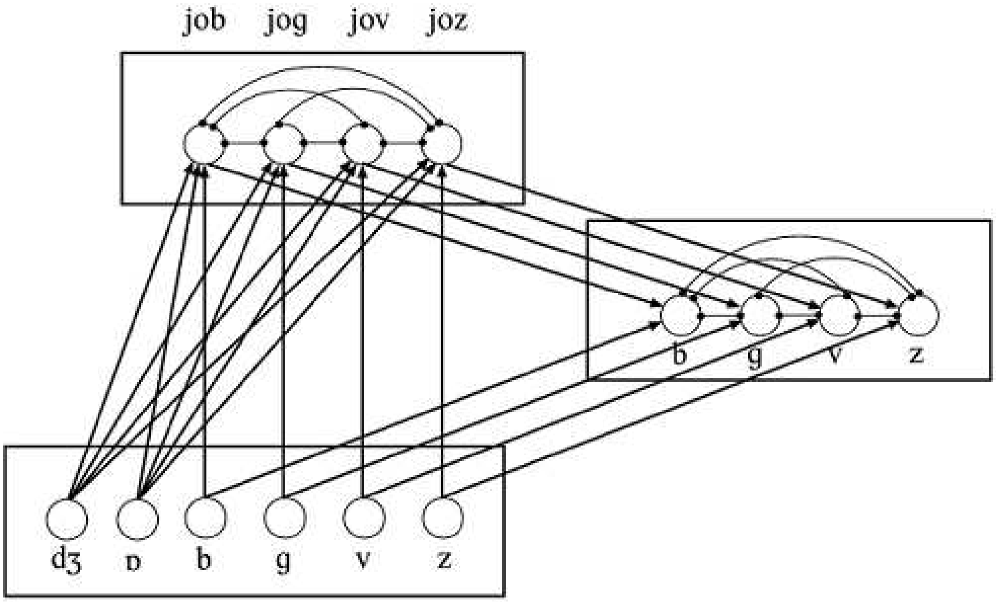

Figure 5 : Le modèle MERGE (Norris et al., 2000). Deux catégories d'informations sont disponibles à la sortie du traitement : la prise de décision (le niveau d'activation des nœuds de la couche de droite) peut aussi bien reposer sur le niveau d'activation des nœuds phonémiques (en bas) que sur celui des nœuds lexicaux (en haut). Le type de représentation utilisé ne permet pas de faire apparaître les boucles de récurrence.

Une question reste cependant posée : celle du statut des unités cachées dans l'émergence des représentations que développe le modèle (cf. Massaro, 1988 pour une critique de la fonction des unités cachées dans les modèles du fonctionnement cognitif). En effet, une Analyse en Composantes Principales effectuée par Norris et al. (2000) a consisté à essayer d'identifier quelles représentations émergent des unités cachées (*Hidden Layer Units*) suite à l'apprentissage. La question est importante puisqu'on pourrait imaginer que ces unités cachées, dont le nom indique qu'elles sont inaccessibles au processus de prise de décision, codent indépendamment les informations lexicales et phonémiques. Certaines unités constitueraient alors des unités lexicales alors que d'autres représenteraient des unités phonémiques. Les phénomènes d'interaction lexique-phonème qui -selon l'interprétation initiale des auteurs- émergeraient en raison de l'existence de procédures d'intégration de ces deux classes d'informations au niveau décisionnel, pourraient alors se produire avant la prise de décision (donc dans le cadre des processus de *codage* de l'information). Ceci reviendrait simplement à implémenter des procédures rétroactives d'une manière différente de celle que l'on trouve dans des modèles d'activation interactive comme TRACE (McClelland & Elman, 1986). L'analyse effectuée par les auteurs montre au contraire que chaque unité cachée code à la fois des informations lexicales et sublexicales. C'est donc la combinaison d'activation de plusieurs unités cachées qui détermine l'activation des unités de sortie (phonémiques ou lexicales). Les auteurs en concluent que rien ne permet de dire qu'il y a eu interaction entre unités au niveau du codage ; c'est au contraire au niveau de l'intégration des informations que se situe l'interaction.

Il faut toutefois noter que, s'il n'y a effectivement pas indépendance des représentations de la couche cachée (*Hidden Layer*) -ce qui implique qu'il n'est pas possible d'expliquer cette *interaction* lexique-phonème par le biais de processus interactifs qui prendraient place entre des unités fonctionnellement *et* structurellement indépendantes- le problème n'est pas pour autant résolu. Si chaque unité cachée a codé simultanément de l'information phonémique *et* de l'information lexicale, ceci les ayant conduit à développer des représentations *hybrides*, ces deux classes d'information ne sont par conséquent pas indépendantes l'une de l'autre. Cette absence d'indépendance est alors bien réelle avant même l'accès à l'étape d'intégration des informations, donc au cours du codage même des diverses représentations. La question qui se pose alors est de savoir si cette caractéristique du modèle est à envisager -en se référant à la typologie de Marr (1982)- comme une particularité *algorithmique* qui serait nécessaire pour être en mesure de simuler des effets d'interaction dans un modèle purement ascendant -auquel cas la non-indépendance des représentations qui se développent dans la couche cachée ne serait qu'une manière particulière de développer dans un programme informatique des processus ascendants aptes à simuler des phénomènes d'interaction, ou si c'est au contraire une caractéristique *computationnelle* qui serait alors à concevoir comme partie intégrante des processus permettant au système simulé, le système cognitif, de développer des effets de ce type. Si l'on admet une interprétation purement algorithmique de la couche cachée, on est en droit d'affirmer qu'un modèle ascendant tel que MERGE peut simuler des phénomènes rétroactifs.

Indépendamment de cette question, il reste que les SRN constituent une classe de modèles particulièrement intéressante dans le cadre du problème qui nous intéresse ici -le rôle des contraintes phonologiques séquentielles dans les processus de segmentation de la parole- car ils sont en mesure d'acquérir des informations sur les régularités d'occurrence des séquences (Elman, 1990) qu'ils reçoivent en entrée au cours de l'apprentissage, ce qui n'est pas le cas dans les modèles IAC comme TRACE (McClelland & Elman, 1986). Or les contraintes phonologiques qui portent sur les régularités des séquences de phonèmes dans la langue pourraient tout à fait être acquises par le biais de processus de cet ordre (Brent, 1996).

2.2. Le recours à des connaissances pré-lexicales

Outre le rôle éventuel des connaissances lexicales dans les processus d'identification phonémique qui, si elles sont aujourd'hui avérées peuvent être modélisées aussi bien par des phénomènes rétroactifs que par des processus d'intégration d'informations, d'autres auteurs ont cherché à étudier le rôle éventuel des connaissances concernant les régularités phonologiques de la langue dans des processus d'identification.

2.2.1. Les données expérimentales

Nous décrivons ici trois groupes d'études qui ont abouti à la conclusion d'une utilisation, par le système cognitif, de connaissances concernant les régularités phonologiques de la langue dans les processus d'identification phonémique. Nous restreignons les études mentionnées à celles qui ont pour objet des contraintes séquentielles portant sur la production de phonèmes adjacents dans la chaîne de parole : (1) la compensation perceptive de la coarticulation (Mann & Repp, 1981), (2) le rôle des contraintes phonotactiques dans la catégorisation de phonèmes ambigus (Massaro & Cohen, 1983), et (3) le rôle des contraintes phonotactiques dans l'identification de phonèmes acoustiquement non-ambigus (Hallé, Segui, Frauenfelder, & Meunier, 1998).

2.2.1.1. La compensation perceptive de la coarticulation

Nous avons déjà présenté rapidement ce résultat, utilisé comme point de départ de l'étude réalisée par Elman & McClelland (1988), dans la partie consacrée au rôle du lexique dans l'identification phonémique (cf. Section 2.1.1.4). Il a été mis en évidence (Mann & Repp, 1981) un rôle des connaissances liées aux contraintes coarticulatoires sur les processus d'identification de phonèmes ambigus. Cet effet se manifeste par une tendance à percevoir un phonème ambigu [^{t}k] comme représentant de la catégorie [t] s'il est prononcé après [ʃ] alors qu'il est plutôt perçu [k] s'il est prononcé après [s]. Ceci s'expliquerait par l'utilisation, de la part des auditeurs, de connaissances concernant le partage du lieu d'articulation dans la production de ces sons de parole, [t] et [s] partageant le trait [+dental] alors que [k] et [ʃ] partagent le trait [+palatal]. La coarticulation induit en effet le système de production de la parole à 'assimiler' certains phonèmes à d'autres. Ici, la production du son dental [s] entraîne les articulateurs à produire le son suivant avec des caractéristiques qui lui sont proches -en l'occurrence le lieu d'articulation- en influençant le contrôle des gestes articulatoires afin de diminuer la distance -exprimée en termes de traits distinctifs- entre ce son et le précédent. Dans la production d'un signal naturel, ces contraintes influencent les caractéristiques acoustiques des sons de parole. Certains sons risquent alors de présenter des caractéristiques qui ne déterminent pas exactement leur représentation phonologique sous-jacente. Le système d'identification phonémique pourrait alors les interpréter comme des phonèmes différents de ceux que souhaitait transmettre le locuteur. Lorsque le système perçoit une séquence de phonèmes qui sont relativement proches, il serait utile de pouvoir compenser cet effet de la coarticulation. On voit donc en quoi ce type de procédures pourrait être utile au processus d'appariement entre forme phonétique de surface et représentation phonologique sous-jacente. L'effet mis en évidence par Mann & Repp (1981) illustre l'aptitude du système de perception de la parole à avoir recours à des connaissances qu'il aurait intégrées concernant les processus de production pour modifier les représentations qui sont dérivées du signal acoustique et parvenir à apparier cette image auditive avec les représentations phonologiques adéquates.

Ces effets mettent en évidence l'utilisation, dans les processus d'identification d'un signal de parole, de contraintes qui régissent le contrôle articulatoire pour la réalisation acoustique des sons. Un ensemble de régularités phonologiques des langues (dites contraintes phonotactiques) intervient également dans les processus de production du signal de parole. Ces contraintes sont dépendantes des langues et intégrées par les locuteurs au cours du processus d'acquisition. Elles définissent un ensemble de 'règles' qui régissent la possibilité (ou l'impossibilité) qu'ont les locuteurs d'une langue donnée de produire certaines séquences de phonèmes dans un signal de parole. Une discussion plus approfondie de la notion de *contraintes phonotactiques* sera proposée dans le Chapitre 4. Cette classe de contraintes constitue le cœur du problème que nous étudions dans le cadre de notre travail. Nous présentons ici les travaux qui ont été réalisés afin d'évaluer le rôle de ces contraintes dans les processus d'identification de la parole sachant que notre étude porte plus spécifiquement sur la question de la *segmentation du signal de parole en mots* ; problématique qui sera abordée dans le chapitre suivant.

2.2.1.2. Le rôle des contraintes phonotactiques dans l'identification de phonèmes ambigus

Si l'on présente à des auditeurs une séquence de parole sans signification dans laquelle l'un des phonèmes est ambigu, ceux-ci ont tendance à donner du son ambigu une interprétation qui préserve les contraintes phonotactiques de la langue. Ce résultat a été mis en évidence par Massaro & Cohen (1983). Dans une tâche de catégorisation phonémique, des locuteurs anglais entendaient des logatomes CCV dans lesquels le phonème médian variait sur un continuum [r]-[l]. Dans chacun des groupes de stimuli, l'une des extrémités du continuum constituait un groupe de consonnes phonotactiquement illégal en début de mot en anglais (par exemple */tla/) alors que l'autre extrémité était parfaitement attestée dans la langue (/tra/) en position initiale de mot. En fonction de la catégorie dans laquelle les participants classaient le phonème ambigu, la séquence de consonnes initiale donnait par conséquent lieu soit à une séquence phonotactiquement illégale dans la langue (*/sra/, */dla/ ou */tla/), soit à une séquence légale (/sla/, /dra/ ou /tra/). Les auteurs observent une tendance des participants à favoriser une interprétation légale de la séquence de phonèmes ; ceci se manifestant par un déplacement de la frontière catégorielle. Dans les séquences /s$^{r}_{l}$a/, les participants ont tendance à percevoir plus souvent [$^{r}_{l}$] comme un représentant de la catégorie /l/ alors qu'ils le perçoivent plutôt /r/ dans /t$^{r}_{l}$a/ et /d$^{r}_{l}$a/. Les auditeurs seraient donc en mesure d'utiliser une certaine catégorie de connaissances ayant trait aux régularités phonologiques de leur langue pour modifier les représentations qui peuvent être construites à partir d'une analyse acoustique du signal de parole.

Pitt (1998) soulève cependant la question, déjà évoquée par Massaro & Cohen (1983), de la cause effective de ce déplacement de la frontière catégorielle dans les résultats précédents. En effet, alors que les séquences */tl/ et */dl/ constituent toutes deux des suites de phonèmes phonotactiquement illégales en anglais, les résultats obtenus par Massaro & Cohen (1983) montrent un déplacement plus important de la frontière catégorielle pour les réponses données aux séquences [t$^{r}_{l}$] par rapport à celles observées pour les stimuli [d$^{r}_{l}$]. Or il apparaît qu'en anglais, la différence de fréquence -en position médiane de mot- entre les séquences /tl/ et /tr/ est plus importante qu'entre les séquences /dl/ et /dr/. Ces effets pourraient alors s'expliquer plus par l'intervention de processus probabilistes que par des *connaissances* fondées sur des règles ou des contraintes distinguant deux catégories de séquences : légales et illégales. Selon cette interprétation, l'émergence d'un déplacement plus important de la frontière pour les paires /tl/-/tr/ que pour les paires /dl/-/dr/ serait lié à un décalage dans les fréquences relatives des groupes de consonnes. Pitt (1998) réplique l'expérience réalisée par Massaro & Cohen (1983) afin d'estimer la contribution relative des sources d'information fréquentielles et phonologiques dans les effets observés. Pour cela, il compare quatre continua dans lesquels les rapports, entre les deux extrémités, en termes de légalité phonotactique ou de fréquence sont en partie orthogonaux. Il compare ainsi les continua utilisés par Massaro & Cohen (1983) à deux autres continua dont les extrémités constituent toutes deux des séquences légales mais présentent une différence dans leurs fréquences d'occurrence (/gr/ vs. /gl/, /gr/ étant plus fréquent -quelle que soit sa position dans les mots- que /gl/) ou pas de différence de fréquence du tout (/br/ vs. /bl/). Cette dernière condition fournit une ligne de base pour estimer la localisation de la frontière entre /r/ et /l/. L'étude d'un éventuel déplacement de la frontière dans le contexte /g/ permet d'estimer le rôle joué par la fréquence indépendamment de la légalité phonotactique. Avec ce contrôle, Pitt (1998) met en évidence un rôle effectif mais faible de la fréquence…

2.2.1.3. Le rôle des contraintes phonotactiques dans l'identification de phonèmes non-ambigus

Abordant une problématique similaire à celle de Massaro & Cohen (1983), Hallé, Segui, Frauenfelder & Meunier (1998) observent une modification dans la perception de séquences illégales *mais* non-ambiguës. Les auteurs comparent, chez des locuteurs français, le traitement de séquences phonotactiquement illégales en début de mot dans la langue (/tl/, /dl/) à leur contrepartie légale (/tr/, /dr/). Contrairement aux stimuli utilisés dans les expériences de catégorisation phonémique (Massaro & Cohen, 1983 ; Pitt, 1998), la consonne liquide médiane des stimuli utilisés n'est pas ambiguë. Par ailleurs, des différences phonétiques caractérisent le phonème /r/, lequel se réalise avec une articulation uvulaire ([ʁ]) en français alors qu'il est alvéolaire ([r]) en anglais. De ce fait, les phonèmes /r/ et /l/ sont, sur le plan acoustique, plus distincts dans la langue française qu'en anglais. Les auteurs prédisent par conséquent un effet plus probable de la légalité sur le percept attaché à la première consonne du groupe qu'à celui de la consonne liquide. Les stimuli utilisés sont des bisyllabes commençant par une séquence de consonnes illégale (/tl/ dans /tlabdo/) ou légale (/tr/ dans /trabdo/). Dans les deux premières expériences, la tâche des sujets consiste respectivement à retranscrire ce qu'ils pensent avoir entendu ou à choisir parmi deux possibilités celle qui leur semble correcte. Les résultats montrent que les sujets tendent effectivement à percevoir les séquences illégales comme des séquences légales. Lorsqu'elle est suivie d'un /l/, l'occlusive dentale initiale (respectivement /t/ et /d/) est dans la majorité des cas perçue comme une occlusive vélaire (respectivement /k/ et /g/). Les participants ne donnent que très rarement une interprétation autre que vélaire (par exemple les bilabiales /p/ et /b/) de la consonne initiale. Dans une tâche de dévoilement progressif (gating, Grosjean, 1985), il apparaît que la consonne initiale est correctement identifiée avant l'occurrence de la liquide (que celle-ci soit un /l/ ou un /r/) ; ce n'est qu'une fois la liquide identifiée que les sujets commencent à privilégier les réponses vélaires par rapport aux réponses dentales lorsque la contrainte phonotactique n'est pas respectée. Les auteurs en concluent que les locuteurs utilisent des connaissances sur les régularités phonotactiques de leur langue pour 'filtrer' le signal acoustique et aboutir à un percept légal dans leur langue.

En reprenant les problèmes posés par l'étude de Massaro & Cohen (1983), on peut à nouveau mettre en doute le rôle de connaissances sur des contraintes permettant de distinguer deux catégories de séquences -légales et illégales- et se demander si des informations probabilistes ne suffiraient pas à rendre compte de ces effets. Du fait, certainement, de leur illégalité phonotactique, les suites dentale + liquide sont non seulement -presque- inexistantes en début de mot mais également, si l'on calcule leur fréquence d'apparition sans prendre en compte leur position dans les mots, beaucoup plus rares que les suites vélaire + liquide (cf. Chapitre 4 pour une analyse distributionnelle des groupes de deux consonnes dans un lexique français). Il est par conséquent possible que cette tendance à percevoir les occlusives dentales comme des vélaires lorsqu'elles sont suivies d'un /l/ puisse s'expliquer par l'utilisation, dans le cadre du système de décodage acoustico-phonétique proprement dit ou au niveau de l'intégration des diverses informations disponibles, de statistiques lexicales sur les probabilités d'occurrence des séquences phonémiques. L'une des difficultés posées par cette interprétation est que les sujets donnent presque exclusivement des réponses vélaires et très peu de réponses bilabiales. Dans le cadre d'une interprétation statistique, ces deux types de séquences ne présentant pas de contraste clair quant à leur fréquence d'occurrence dans le lexique, devraient se retrouver à part égale dans les réponses des sujets. Or on observe au contraire une dissymétrie marquée en faveur des réponses vélaires. L'interprétation phonotactique ne fournit cependant pas une solution plus claire à cette dissymétrie. Les séquences /kl/ et /pl/ sont tout aussi légales l'une que l'autre. Il n'existe par conséquent pas de justification phonotactique au déséquilibre observé. Par contre, on peut tout à fait envisager une proximité acoustique plus importante (qui pourrait se mesurer à l'aide d'une matrice de confusions dans une tâche perceptive) entre /t/ et /k/ qu'entre /t/ et /p/ ; ceci pourrait expliquer la forme des résultats et s'appliquer aussi bien dans le cadre d'une interprétation phonotactique que statistique du biais observé dans les expériences de Hallé et al. (1998).

Résumé

Nous avons vu, dans la première partie de ce chapitre, les diverses difficultés posées par la modélisation de l'appariement entre un signal

acoustique et des représentations linguistiques stockées en mémoire. Dans la seconde partie, nous nous sommes attaché à présenter un certain nombre d'études dans lesquelles la question de l'utilisation de connaissances de haut niveau (lexicales ou phonologiques) pour influencer le codage de l'information auditive s'est posée. Il apparaît clairement que la réponse à cette question n'est pas disponible à l'heure actuelle. Seules des convictions personnelles permettent aujourd'hui de se situer d'un côté ou de l'autre de la 'frontière théorique'. Cette analyse des données disponibles nous a cependant permis d'évoquer la difficulté que l'on peut rencontrer à tester certaines hypothèses en rapport avec le traitement du langage, difficulté en grande partie associée au lien étroit qui peut exister entre des variables telles que la densité de voisinage lexical et la fréquence des diphones (rôle du lexique) ou la légalité phonotactique et la fréquence (rôle des contraintes phonotactiques). Dans le chapitre suivant, nous abordons de manière plus précise la question centrale de notre travail : le rôle éventuel des contraintes phonologiques séquentielles de la langue dans la segmentation du signal de parole en mots en présentant deux approches complémentaires des processus de segmentation lexicale.

Chapitre 2

Segmentation du signal de parole en mots

SEGMENTATION DU SIGNAL DE PAROLE EN MOTS

Il peut sembler étonnant de poser le problème des procédures mises en œuvre pour découvrir les frontières qui séparent les mots dans un message parlé comme un phénomène qui serait en partie indépendant des processus de reconnaissance des mots. Supposons en effet que les difficultés concernant la variabilité -spectrale et temporelle- du signal de parole aient été résolues, et que l'on soit aujourd'hui en mesure de proposer un modèle valide et non-réfuté de l'appariement entre image auditive et représentations phonétiques. Il serait alors possible de mettre au point un système artificiel capable d'identifier avec certitude la chaîne de phonèmes correspondant au signal acoustique qui lui est fourni en entrée. Supposons maintenant que ce système dispose également de procédures lui permettant de traduire la chaîne de phonèmes en représentations phonologiques. Ce système disposerait alors de représentations sous-jacentes directement appariables avec les éléments stockés dans son lexique. On peut penser qu'il suffirait alors d'envisager, à partir de cette chaîne d'entités phonologiques sous-jacentes, l'ensemble des suites de mots possibles pour retrouver la séquence de mots effectivement produite par le locuteur. De manière simplifiée, c'est l'une des stratégies qui ont été proposées dans plusieurs modèles de la segmentation lexicale en psycholinguistique. Cette procédure est aussi largement utilisée en Reconnaissance Automatique de la Parole (RAP). La segmentation du signal de parole en mots se fait alors par comparaison de la séquence phonémique avec les mots qui sont stockés dans le lexique. La séquence la plus plausible est sélectionnée ; ce qui permet d'aboutir à une segmentation relativement efficace du signal de parole en mots. Nous verrons, dans ce chapitre, que cette stratégie n'est peut-être pas la mieux adaptée aux processus de traitement de la parole. Pour ne mentionner qu'un argument, on notera le rôle éventuel des connaissances de haut niveau dans les processus d'identification phonémique, notamment le rôle possible de l'identification lexicale dans l'appariement entre image auditive et représentations linguistiques. Si les connaissances lexicales ont un rôle à jouer dans l'identification phonémique, il est important d'étudier la segmentation en mots comme un problème en soi ; des procédures aptes à guider la localisation des frontières lexicales pourraient ainsi guider l'appariement entre représentation auditive du signal acoustique et formes phonologiques abstraites en facilitant la reconnaissance des mots. L'objet de ce chapitre est à la fois de présenter divers modèles de la segmentation du signal de parole en mots et -parallèlement- d'illustrer les difficultés spécifiques que pose l'étude de ces processus. Nous décrivons en premier lieu les modèles qui reposent sur des processus de sélection ou de compétition lexicale pour présenter des positions alternatives dans lesquelles est intégrée l'utilisation d'informations non-lexicales (métriques, phonotactiques, statistiques).

1. La segmentation comme conséquence de l'accès au lexique

Depuis quelques années, on envisage que l'une des sources essentielles d'information pour la localisation des frontières de mots dans un signal de parole est l'identification même des mots qui le constituent. Certains modèles d'accès au lexique se sont alors présentés comme un choix particulièrement bien adapté au problème de la segmentation lexicale d'un signal de parole continue. Ces modèles reposent essentiellement sur le principe de la *sélection progressive des candidats lexicaux* les mieux appariés avec l'entrée acoustique.

1.1. Cohort

Le modèle COHORT (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978) est le premier à introduire la notion de reconnaissance d'un mot (ou d'accès au lexique) comme un processus de sélection progressive parmi un ensemble de candidats lexicaux possibles. Ce modèle comporte 2 étapes : (1) une étape d'*activation* des candidats lexicaux appariés avec le début du signal suivie de (2) une étape de *dé-activation* des candidats dont la structure phonétique dévie du signal. Il repose sur l'une des caractéristiques essentielles du signal de parole : son déroulement dans le temps. Contrairement à l'identification de mots écrits, situation dans laquelle le système visuel est en mesure de coder intégralement -pour des mots de longueur moyenne- l'information pertinente avec une seule fixation visuelle, la reconnaissance des mots parlés implique un *codage progressif* de l'information. Marslen-Wilson & Welsch (1978) utilisent cette particularité des stimuli acoustiques pour proposer un modèle spécifique de la reconnaissance des mots dans la modalité auditive. La première étape consiste à déclencher l'activation d'une *cohorte* de mots possibles à partir de l'information disponible au début du signal acoustique (les 150 premières millisecondes, ce qui correspond approximativement à un ou deux phonèmes). A partir de cet ensemble de mots possibles, le système exclut les candidats (dans COHORT I) ou réduit progressivement leur activation (dans COHORT II) quand ils ne sont plus appariés avec l'entrée acoustique. Ce processus se poursuit jusqu'à aboutir à un choix lexical unique. Le point qui correspond à l'instant auquel le signal acoustique fournit une information qui permet d'identifier sans ambiguïté le mot prononcé est appelé *Point d'Unicité* (PU) ou *Point de Reconnaissance* (PR)[11]. Nous ne nous attarderons pas ici sur les problèmes liés aux conditions d'exclusion des candidats de la cohorte (doit-on supprimer les candidats qui ne sont pas appariés avec l'entrée acoustique ou simplement diminuer leur niveau d'activation ? -COHORT I vs. COHORT II), pas plus que sur la question de l'existence de processus d'inhibition ascendante ou de rétroaction du lexique sur les représentations phonémiques. On consultera notamment Frauenfelder (1996) pour des discussions relatives à ces problématiques. Notre objectif est ici de définir les caractéristiques fondamentales du modèle en rapport avec le problème de la segmentation d'un signal de parole continue en mots afin de fournir les bases d'une analyse des données présentées comme reflétant le recours à des connaissances sur les régularités phonotactiques de la langue.

1.1.1. Prédiction des frontières lexicales avant la fin acoustique du mot

Le modèle COHORT (I, Marslen-Wilson & Welsh, 1978; ou II, Marslen-Wilson, 1987) apporte une solution au problème de la localisation des frontières lexicales dans un signal de parole continue par le biais du processus même d'accès au lexique. Pour cela, il permet d'identifier un mot avant sa fin acoustique. Or, si l'on est en mesure de reconnaître un mot avant qu'il ait été traité dans son intégralité, on est en mesure de savoir à l'avance où il va se terminer et, par conséquent, où commencera le mot suivant. Ainsi, le verbe /apɛrsəvwar/ ('apercevoir') peut être isolé comme unique dans le lexique français dès l'occurrence du phonème /ə/.

[11] Selon que l'on considère le point à partir duquel le mot est unique dans un dictionnaire de la langue (PU) ou celui à partir duquel les locuteurs l'identifient avec certitude (PR ; ces deux 'points' peuvent notamment différer en raison de la fréquence d'usage des mots).

Lorsqu'en modalité auditive on a traité le flux acoustique qui correspond à la séquence de phonèmes /apɛrsə/, il n'existe pas dans le lexique d'autre possibilité que ce verbe. On est donc en mesure, dès cet instant là, de prédire que la fin du mot en cours correspondra à la séquence /vwar/ et que ce mot se terminera avec le phonème /r/. On peut alors prévoir avec certitude que le mot suivant commencera immédiatement après le phonème /r/. Il est par conséquent possible, grâce à des processus de reconnaissance des mots, de localiser les frontières lexicales avant qu'elles apparaissent effectivement.

L'aptitude du modèle COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987) à déclencher le processus d'accès lexical en parole continue repose essentiellement sur sa capacité à localiser le début des mots -on parle de modèle à *alignement initial*. En effet, c'est l'information acoustique fournie par le début du signal acoustique qui permet d'engendrer un ensemble d'hypothèses lexicales sur lesquelles vont s'opérer les processus de sélection. Seuls les mots dont le début est apparié avec l'entrée acoustique sont intégrés dans la cohorte. Ce modèle nécessite donc, dans le cas du traitement d'un signal de parole continue, une prédiction correcte de la prochaine frontière lexicale pour être en mesure de générer une cohorte pertinente pour le mot suivant. Si le système omet une frontière lexicale, il n'est plus en mesure de générer en temps réel la cohorte initiale de candidats lexicaux appariés avec l'entrée. Tout retard rencontré dans la localisation de la frontière lexicale induit alors une impossibilité -au moins transitoire- de générer une cohorte correspondant au début du mot suivant. Dans cette situation, seule une procédure de récupération par un retour en arrière dans les processus de traitement permettrait de récupérer le cours normal du traitement. Il n'est pas envisagé de pouvoir générer un ensemble de candidats qui seraient appariés avec le signal autrement que par leur début. Ce modèle est donc efficace uniquement s'il existe dans le lexique une grande quantité de mots -sinon tous- qui peuvent effectivement être reconnus avant leur fin acoustique. Supposons que l'on traite par exemple la suite /elef/, cette séquence de phonèmes permet de restreindre la cohorte à quelques mots du lexique français : 'éléphant', 'éléphanteau', 'éléphantiasis'. L'existence de ces deux derniers empêche le système d'accès au lexique d'être en mesure d'affirmer -dès l'occurrence du phonème /f/- que le prochain mot dans la séquence de parole commencera après le son /ɑ̃/. Il pourrait tout aussi bien commencer après le son /o/ si le mot en cours de traitement est en réalité 'éléphanteau'. On objectera que ceci ne met pas en échec le processus de localisation des frontières lexicales à partir de l'accès au lexique. Effectivement, si le mot devient unique sur son dernier phonème, il permet d'insérer une frontière lexicale et de relancer un nouveau processus d'accès au lexique à partir de ce point. Cependant, cette objection n'est valable que si c'est effectivement le mot ‘éléphanteau’ qui est présenté. Si le mot à identifier est en fait ‘éléphant’ et que la séquence de phonèmes subséquente commence par /g/ comme dans ‘un éléphant gris’, c’est l’occurrence du phonème /g/ qui conduira le système à identifier le mot ‘éléphant’ (puisqu’il n’y a pas dans le lexique de mot commençant par /elefɑ̃g/). Le système devra alors revenir sur ses prédictions pour générer la cohorte pertinente à partir du son /g/. COHORT II (Marslen-Wilson, 1987) apporte en partie une solution à ce problème en introduisant la notion de niveau d’activation et en faisant appel à la fréquence des mots pour moduler l’activation des candidats. Le processus de sélection de la réponse appropriée (qui est alors assimilable à un processus intégratif) prend appui sur la fréquence relative des différents candidats activés et favorise les plus fréquents. Dans l’exemple précédent, ‘éléphant’ est beaucoup plus fréquent que les mots ‘éléphanteau’ et ‘éléphantiasis’. Il aura donc plus tendance à se maintenir dans la cohorte ; ce qui évitera peut-être un rejet trop précoce si le seuil de reconnaissance est bien fixé. On peut alors envisager que le modèle déclenche l’activation d’une cohorte sur le phonème /g/ avant même de pourvoir prédire que c’est effectivement le bon choix avec 100% de certitude.

1.1.2. Point d’Unicité tardif et génération d’une cohorte

Les difficultés inhérentes à ce modèle s’accentuent donc s’il existe dans le lexique des mots qui, même au niveau de leur fin acoustique, ne sont toujours pas uniques. Dans l’analyse d’un lexique anglais informatisé (Kucera & Francis, 1967), Luce (1986) montre que la quantité des mots qui ne sont pas uniques sur leur phonème final est considérable (74% des mots de 3 phonèmes et 36% des mots de 4 phonèmes ; cf. Tableau 1). Ce phénomène est d’autant plus manifeste que les mots sont courts. Frauenfelder & Peeters (1990), dans une analyse similaire effectuée sur un échantillon de mots néerlandais provenant de la base de données CELEX, obtiennent des données tout à fait comparables (61% de mots à PU tardif pour 3 phonèmes, 31% pour 4 phonèmes, 15% pour 5 phonèmes). Une stratégie consistant à prédire les frontières lexicales par le biais de processus de sélection reposant sur un alignement initial ne serait donc réellement efficace que pour le traitement des mots longs. Dans le cas des mots courts, le système devrait soit postuler des frontières lexicales en permanence afin de ne pas en omettre une seule, soit mettre en place un processus de retour en arrière lorsqu’il s’aperçoit qu’il en a laissé passer une (ceci même s’il incorpore un paramètre fréquentiel pour maintenir les mots fréquents -qui sont aussi les plus courts- dans la cohorte initiale).

Tableau 1 : Pourcentage de mots dans le lexique anglais dont le Point d'Unicité se situe après le phonème final. Extrait des résultats de Luce (1986) pour les mots comptant entre 3 et 7 phonèmes.

Nb. de phonèmes	Nb. de mots dans l'échantillon	Pourcentage de mots
3	1839	74.17
4	3025	35.64
5	3172	16.36
6	3063	8.56
7	2735	5.59
8	2210	4.75

L'existence d'un délai entre la fin acoustique des mots à PU tardif et leur reconnaissance est mise en évidence par Grosjean (1985). Il utilise une tâche de dévoilement progressif (*Gating*) dans laquelle des auditeurs écoutent des phrases contenant une quantité d'information acoustique variable. Les stimuli que reçoivent les participants vont progressivement des premières périodes acoustiques à l'intégralité du signal (cf. Figure 6 pour un exemple des stimuli utilisés dans une tâche de gating). Grosjean (1985) montre que les mots courts ayant des prolongements possibles dans le lexique ne sont identifiés avec un taux de certitude élevé que bien après leur fin acoustique effective, alors même que le mot suivant est déjà en cours de traitement. C'est donc l'information acoustique ultérieure qui permet, pour des mots à PU tardif présentés dans un signal de parole continue, d'accéder à l'identification du mot précédent. Dans le cas des mots courts, il est donc la plupart du temps impossible de prédire à l'avance la localisation de la frontière lexicale, même s'ils sont en moyenne plus fréquents que des mots longs. Tant que le système est capable de localiser une frontière lexicale avant de commencer à traiter le mot suivant, le processus est efficace. Cependant, du fait de la quantité importante de mots à PU tardif dans la langue, il est peu probable que cette procédure puisse être fiable dans une situation naturelle de reconnaissance des mots en parole continue. Il faut toutefois noter l'intérêt d'un modèle comme COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987) qui, malgré les difficultés liées à un alignement sélectif sur les débuts de mots, fournit une représentation des processus d'accès au lexique spécifique de la modalité auditive en prenant en considération l'aspect sériel des processus de reconnaissance des mots parlés. Or plusieurs résultats expérimentaux montrent qu'il est important de tenir compte de cette caractéristique du signal de parole (Mattys, 1997).

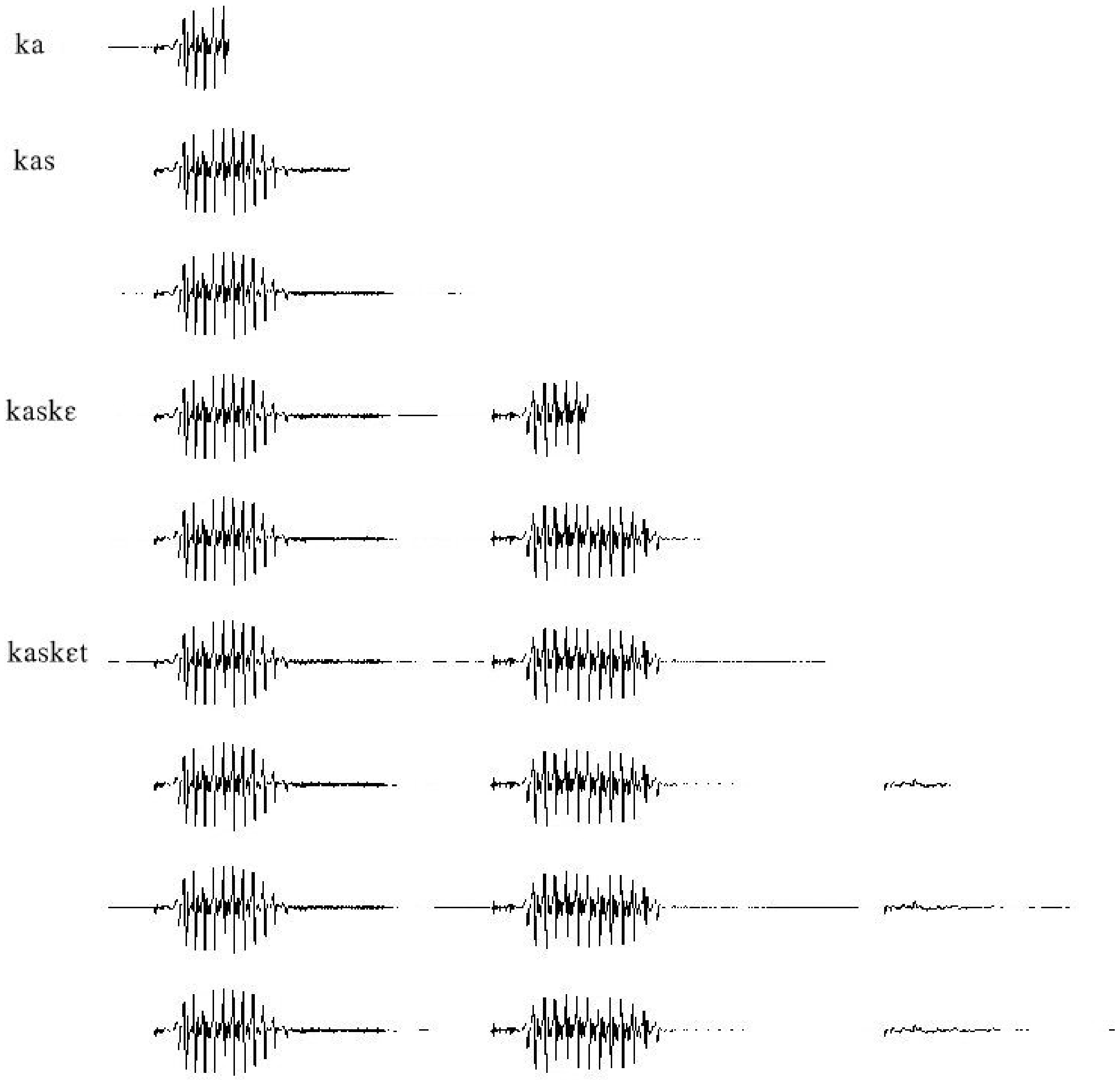

Figure 6 : Exemples de stimuli utilisés dans une tâche de dévoilement progressif (ou *gating*). Ici, le mot utilisé pour créer une série est 'casquette'. La colonne de gauche indique approximativement les changements de phonèmes perçus.

1.2. Trace

Le modèle TRACE (McClelland & Elman, 1986) s'inspire en grande partie des propositions introduites dans COHORT (et notamment Cohort II, Marslen-Wilson, 1987). Il reprend le principe de sélection progressive des candidats lexicaux à partir de l'accumulation des informations fournies en entrée et implémente le processus de segmentation lexicale comme une conséquence des processus d'identification des mots. La notion de niveau d'activation telle qu'elle est proposée dans COHORT II (Marslen-Wilson, 1987) est centrale dans TRACE (McClelland & Elman, 1986). Il fournit cependant une solution au problème de la segmentation lexicale d'un signal de parole constitué de mots à Point d'Unicité tardif en introduisant les notions d'*alignement multiple* et de *compétitions lexicales*.

1.2.1. Alignement exhaustif des activations lexicales

Contrairement au principe d'alignement initial des hypothèses lexicales implémenté dans les deux versions de COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987), McClelland & Elman (1986) introduisent une procédure d'*alignement multiple* (ou *exhaustif*). Dans le modèle COHORT, les hypothèses lexicales (la cohorte de candidats) sont générées uniquement à partir des information acoustiques qui sont considérées comme des débuts de mots. Si le début du signal acoustique correspond au phonème /v/, les éléments lexicaux activés sont ceux qui commencent par /v/ dans le lexique. La séquence de phonèmes /vag/ donnera lieu, dans une première phase, à l'activation de mots comme 'varan', 'vigne', … qui commencent tous par le phonème /v/ (ce qui correspond à 864 mots dans la base de données lexicale Brulex, Content, Mousty, & Radeau, 1990). Un phonème devant nécessairement être considéré comme un début de mot pour générer l'activation d'une cohorte, aucun autre ensemble de candidats ne sera généré tant que le mot commençant par /v/ n'aura pas été identifié, permettant ainsi de déterminer le début du mot suivant et, par conséquent, le phonème qui devra donner lieu à la prochaine cohorte de candidats.

Dans TRACE, au contraire, les hypothèses lexicales peuvent être générées à partir de n'importe quel point du signal. Il n'est pas nécessaire de considérer une partie du signal comme un début de mot pour engendrer une cohorte de candidats. Ainsi, dans le mot /vag/ ('vague'), chaque portion temporelle du signal -de manière simplifiée, chaque phonème[12]- donnera lieu à l'activation d'un ensemble d'hypothèses lexicales contenant ce son. Chaque phonème va donc donner lieu à un ensemble d'hypothèses lexicales dont l'activation ne dépend pas de la localisation préalable d'une frontière entre mots. Ici, une quantité de mots bien plus importante sera activée en raison de cet alignement exhaustif (dans la base de données Brulex, Content et al., 1990, 864 mots commencent par /v/, 2017 par /a/ et 844 par /g/). L'activation des candidats indépendamment d'hypothèses sur les débuts de mots permet de résoudre le problème des mots à PU tardif en faisant reposer la segmentation lexicale sur des phénomènes de compétition entre les candidats activés.

[12] En réalité, et afin de simuler les effets de la coarticulation, le modèle génère un ensemble de candidats tous les 3 cycles de traitement ; ce qui donne donc lieu à plusieurs réseaux lexicaux pour chaque phonème.

1.2.2. Compétitions entre candidats lexicaux

L'utilité d'un alignement exhaustif dans la génération des hypothèses lexicales est intimement liée à la mise en œuvre de phénomènes de compétition entre les candidats lexicaux activés. On entendra ici le mot 'compétitions' dans son acception connexionniste, c'est à dire en termes de régulation du niveau d'activation des différents candidats lexicaux entre eux. Il y a, dans COHORT II (Marslen-Wilson, 1987), un phénomène similaire à ce que l'on peut appeler 'compétitions' : à partir d'une cohorte de candidats lexicaux, le système fait partiellement reposer la sélection finale du candidat sur la fréquence relative des mots qui sont maintenus dans la cohorte, mais à aucun moment les unités lexicales activées ne sont censées modifier le niveau d'activation des autres unités. Cette procédure de pondération des activations en fonction de la fréquence des mots correspond en fait à des processus intégratifs ou décisionnels mais pas à des phénomènes de compétitions lexicales tels qu'ils sont envisagés dans le courant connexionniste. On parlera ici de *sélection lexicale*. La sélection lexicale ne trouve sa source que dans les informations fournies par le flux ascendant d'informations et éventuellement dans les processus de modulation du niveau d'activation des candidats à partir de leur fréquence relative. On réserve l'appellation de *compétitions lexicales* à des modèles dans lesquels les unités lexicales s'influencent mutuellement. Cette influence mutuelle se manifeste par des processus d'interaction entre les représentations lexicales activées (par exemple, inhibitions latérales entre unités lexicales). Ces deux composantes du modèle TRACE (alignement exhaustif et compétitions lexicales) sont désignées comme des processus d'activation et de compétition interactives (IAC, *Interactive Activation and Competition*) qui permettent la propagation des informations à travers les diverses unités (traits acoustiques, phonèmes, mots). Cette propagation de l'activation permet une interaction globale entre les unités et aboutit après un certain délai à une stabilisation de l'état du réseau.

Afin de simuler le décours temporel du traitement de l'information, on laisse s'écouler entre chaque entrée un certain nombre de cycles de traitement et, pour être en mesure de prendre en compte les aspects sériels de la succession des phonèmes dans le signal, on réplique le réseau après un nombre arbitraire de cycles. Après chaque cycle, le niveau d'activation des unités du réseau est enregistré. Une fois qu'un état relativement stable est atteint, on est en mesure d'estimer le comportement du réseau et de savoir quels mots (quels phonèmes, quels traits acoustiques) sont les plus activés à un moment donné du traitement et, par conséquent, quelles sont les frontières lexicales prédites.

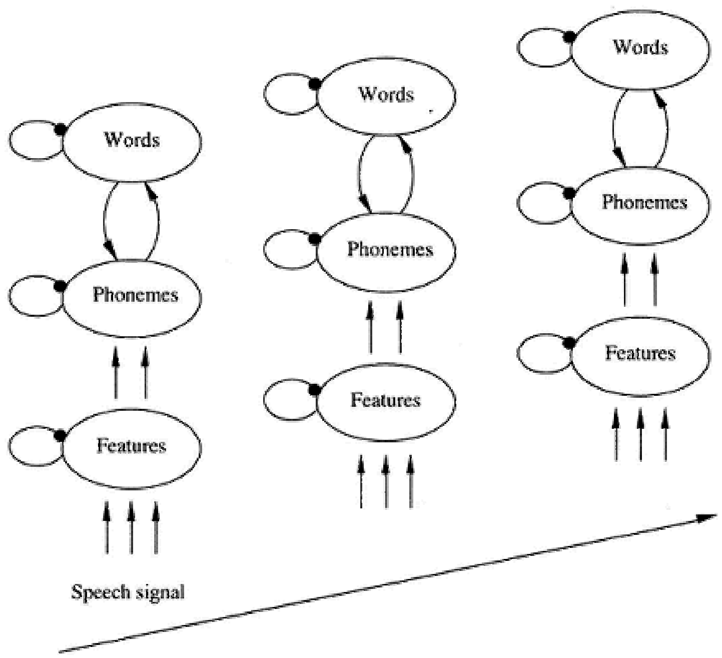

Figure 7 : Représentation graphique du réseau implémenté dans TRACE (McClelland & Elman, 1986). Extrait de Frauenfelder (1996).

1.2.3. Réalisme des procédures implémentées dans TRACE

Le modèle TRACE permet donc de résoudre les difficultés posées par les mots à Point d'Unicité tardif puisque leur reconnaissance ne repose pas sur l'activation d'une cohorte de candidats alignés avec le début de la séquence acoustique mais sur des compétitions entre candidats activés selon une procédure d'alignement exhaustif. Certaines contraintes incitent cependant à poser la question de l'efficacité de ces processus pour l'identification d'une chaîne de mots à partir d'une séquence de phonèmes. L'implémentation des méthodes de représentation sérielle du signal de parole semble par exemple peu adaptée au traitement de séquences de parole relativement longues si le modèle devait utiliser un lexique de taille réaliste. En effet, pour représenter l'ordre des segments dans la séquence de phonèmes, chaque début possible de mot correspond à une réplication du réseau précédent. Les divers réseaux générés sont reliés entre eux afin d'aboutir finalement à une représentation de la distribution des activations en fonction du temps. Plus la séquence à traiter est longue, plus le nombre de réseaux générés par TRACE (McClelland & Elman, 1986) est important. Si l'on souhaite utiliser un lexique contenant autant de mots qu'un locuteur adulte moyen (cinquante à cent mille mots), l'augmentation de la longueur de la séquence à traiter induit un accroissement colossal de la complexité du réseau. Bien que ce paramètre puisse être considéré comme un simple problème algorithmique (cf. la typologie de Marr, 1982) qui ne concernerait que la méthode informatique utilisée pour refléter les capacités du modèle à coder des événements sériels, il serait intéressant d'être en mesure de traiter ces aspects du signal de parole avec un algorithme plus économique, ne serait-ce que pour alléger la procédure de simulation des données comportementales. Un modèle comme TRACE (McClelland & Elman, 1986) pose cependant des problèmes qui ne se réduisent pas uniquement à un ‘manque d’élégance’ algorithmique mais qui reposent aussi sur ses aspects computationnels -c’est à dire la théorie du fonctionnement supposé être effectivement mis en œuvre par le système simulé. On remarquera notamment la quantité importante d’homophones dans les langues (surtout en français). Dans la représentation phonémique de la phrase /nuzavjɔ̃kaʃtelɑ̃vlɔpsyrləʃɑ̃/ (‘Nous avions cacheté l’enveloppe sur le champ’), El-Bèze (1996) trouve, dans un réseau d’homophones, plus de 3 millions de ‘chemins’ possibles allant chacun du début à la fin de la phrase[13]. Il est donc coûteux -en termes de puissance de calcul- de retrouver, dans une séquence de phonèmes de longueur moyenne, la suite des mots qui sont effectivement présents. L’obstacle majeur de ce type d’approche est donc en partie lié à la puissance de calcul nécessaire à implémenter cette stratégie dans un système artificiel de perception de la parole. En effet, pour aboutir avec certitude à la séquence de mots qui correspond à l’entrée acoustique, il faut garder une trace mnésique de cette entrée pendant une durée assez longue. Evidemment, il est difficile de comparer la puissance de calcul d’un ordinateur et celle d’un cerveau dans le choix d’un modèle adéquat du fonctionnement cognitif. La question n’est pas de parvenir, dans le domaine de la recherche fondamentale, à développer des modèles informatiques qui seraient en mesure de traiter le signal aussi rapidement que le système cognitif humain avec des ordinateurs qui fonctionnent rarement avec plusieurs processeurs en parallèle. Les critères à adopter doivent au contraire nous conduire à développer des modèles qui reflètent les processus effectivement mis en œuvre par ce système. On peut néanmoins supposer que le système humain a développé des processus cognitifs économiques qui lui permettent de réaliser les tâches avec le moins de complexité possible. On connaît notamment les limites du système de la mémoire de travail. Or la mise en œuvre d’un système de segmentation lexicale qui reposerait uniquement sur des processus de compétitions nécessiterait le recours à une mise en mémoire parfois vertigineuse d’une partie des traitements qui ont précédé. Ceci induit à remettre en question la validité d’une approche consistant à faire reposer presque exclusivement la localisation des frontières de mots sur les processus d’accès au lexique.

Il est cependant clair que les processus de compétition lexicale sont partie intégrante du système d’accès au lexique. Plusieurs travaux mettent ainsi en évidence des effets du nombre de

[13] Sans prendre en compte la plausibilité syntaxique ou sémantique de la séquence qui en découle. Ce nombre peut sembler énorme. Il faut cependant noter qu’il prend en compte les variantes de prononciation. On peut penser que, même sans ces variantes, on aboutirait à une quantité considérable de ‘chemins’ possibles.

compétiteurs lexicaux ou de l'existence de compétiteurs plus fréquents dans les latences de reconnaissance des mots (McQueen, Norris, & Cutler, 1994; Norris, McQueen, & Cutler, 1995). Il pourrait être intéressant de diminuer la puissance de calcul nécessaire à la mise en compétition d'un nombre aussi considérable d'éléments du lexique sans pour autant supprimer cette particularité du système.

1.3. Shortlist

Pour aboutir à une segmentation lexicale, le modèle SHORTLIST (Norris, 1994) offre des procédures de traitement à la fois plus économiques et réalistes que celles implémentées par McClelland & Elman (1986) dans TRACE. Cette économie dans la puissance de calcul repose sur deux différences : le recours à des boucles de récurrence (Jordan, 1986; Elman, 1990 ; cf. aussi Chapitre 1, Section 2.1.2.2) et la restriction du nombre de candidats activés. Norris, McQueen, & Cutler (1995) ont également implémenté des procédures prélexicales fondées sur une classe de régularités phonologiques -les alternances accentuelles- afin d'influencer les activations lexicales et d'accélérer le processus de sélection des candidats appropriés.

1.3.1. Représentation de l'ordre par récurrence

Au moment de la publication de l'article qui présentait le modèle TRACE (McClelland & Elman, 1986), il n'existait pas encore de procédure économique pour traiter les aspects temporels et / ou sériels de l'entrée d'un modèle. Nous avons vu dans la partie consacrée à TRACE qu'une réplication des réseaux d'activation avait été mise en œuvre pour coder l'ordre des segments phonémiques et leur position dans les mots. Cette méthode induit une croissance exponentielle de la charge de traitement lorsqu'on augmente simultanément la taille de la séquence et celle du lexique. La même année, Jordan (1986) publie un rapport présentant la notion de récurrence locale dans un réseau de neurones. Cette procédure consiste à renvoyer aux unités d'entrée l'information reçue au cours du cycle de traitement précédent par le biais d'une couche d'unités supplémentaire. L'état du système au temps `t` est alors en partie déterminé par son état au temps `t-1`. De fait, lorsque les unités intermédiaires reçoivent l'information fournie par les unités d'entrée, elles disposent également d'une 'mémoire' du cycle de traitement qui a précédé (cf. Figure 8). La méthode de représentation de la structure temporelle d'une séquence par le biais du principe de récurrence est présentée dans Elman (1990). Cette procédure peut être implémentée dans un Réseau Ascendant (*Feedforward Network*) qui prend alors le nom de Réseau Récurrent Simple (*Simple Recurrent Network* -SRN-, cf. MERGE, Chapitre 1). Cette méthode de récurrence de l'information avec insertion d'un délai fournit une mémoire temporelle au modèle qui lui permet de coder à la fois les durées et l'ordre des phonèmes ainsi que leur position dans les mots en activant un seul réseau lexical. Dans SHORTLIST, cette procédure est *simulée* par une recherche classique dans un dictionnaire afin de générer un ensemble d'hypothèses lexicales unique pour l'ensemble des points du signal[14]. Les SRN présentent cependant des propriétés spécifiques qui n'en font pas des candidats adéquats pour simuler les processus de reconnaissance des mots.

1.3.2. Réseaux récurrents simples et enchâssement lexical

Si un réseau récurrent simple est entraîné à reconnaître 'car' et 'cargo' au cours de la phase d'apprentissage, la présentation de la séquence /kargo/ en test lui permettra dans une première étape de reconnaître 'car'. Lorsque /go/ est présenté, le niveau d'activation de l'unité correspondant au mot 'cargo' augmente mais le SRN n'est pas conçu pour réduire l'activation de 'car'. Ce type de modèle est uniquement capable de donner une 'interprétation' de la séquence au moment où il reçoit l'entrée. Il ne peut pas revenir en arrière pour réduire le niveau d'activation de 'car' à partir de l'information acoustique subséquente dans le but de sélectionner l'interprétation la mieux adaptée. Si l'on imagine que les niveaux d'activation reflètent la réponse qui serait donnée par un auditeur, le modèle *identifie* d'abord 'car' puis, avec le traitement du contexte subséquent, 'cargo' ; mais à aucun moment il ne considère 'car' comme faisant partie intégrante du mot 'cargo'. Malgré la possibilité qu'ont les mots enchâssés d'être

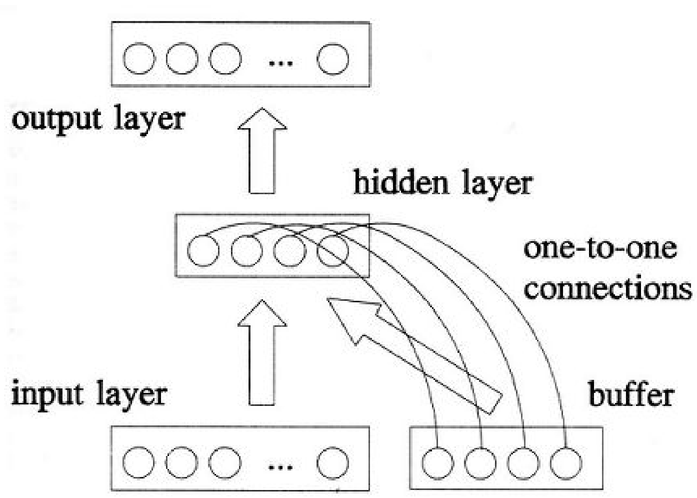

Figure 8 : Illustration d'une boucle de récurrence locale dans un réseau neuronal (extrait de Murre & Goebel, 1996).

[14] On voit ici que l'enjeu n'est pas tant de fournir des procédures économiques de traitement (le codage de l'ordre dans un SRN est plus économique que celui implémenté dans TRACE mais la procédure d'apprentissage pourrait durer plusieurs semaines avec un lexique de taille moyenne) que d'implémenter une représentation réaliste du codage temporel et sériel des événements dans un modèle du système cognitif.

transitoirement activés (Isel & Bacri, 1999 ; Vroomen & De Gelder, 1997), le système de traitement de la parole aboutit toujours à un choix pour le découpage lexical et l'identification des mots présents dans l'énoncé. Avec un SRN, tout se passe comme si nous percevions dans la chaîne parlée toutes les combinaisons possibles de mots sans être en mesure de choisir la bonne. Ces réseaux constituent une classe de modèles bien adaptée à l'identification phonémique car un phonème ne peut pas en contenir un autre. Dans le cas de la reconnaissance des mots, ils doivent inclure des procédures spécifiques qui les conduisent à considérer chaque partie de l'entrée comme provenant d'un mot unique. Dans SHORTLIST, le résultat de ce qui devrait être simulé par un SRN (mais qui repose en fait sur une recherche classique dans un lexique informatisé) est transféré vers un réseau d'Activation et de Compétition Interactive similaire à TRACE. Le nombre de candidats lexicaux qui sont envoyés au réseau IAC est cependant restreint par une limitation arbitraire du nombre de compétiteurs (c'est la shortlist, qui est composée, dans les simulations présentées par Norris, 1994, de 3 à 30 candidats). Ils entrent alors en compétition les uns avec les autres dans un réseau lexical identique au niveau lexical que l'on trouve dans TRACE de manière à aboutir à l'identification d'une unique séquence de mots. Cette restriction de la taille de l'ensemble des hypothèses lexicales et du nombre de réseaux générés en parallèle -qui se réduit à un réseau unique- permet d'utiliser un lexique bien plus grand (environ 50 000 mots) dans les simulations et de conserver la mise en œuvre des phénomènes de compétition entre candidats lexicaux comme une caractéristique essentielle des processus de segmentation de la parole en mots.

1.3.3. Implémentation de procédures de segmentation prélexicales

Plus récemment, des procédures de segmentation prélexicales ont été implémentées dans ce modèle afin de guider la reconnaissance des mots à partir de règles dérivées des régularités phonologiques de l'anglais. En effet, des données comportementales ont mis en évidence le recours à des informations prélexicales pour segmenter le signal de parole en mots. Des connaissances sur certaines régularités des langues (phonologiques ou distributionnelles) permettraient de guider les processus de segmentation de la parole et pourraient faciliter la tâche du système. Ces effets sont simulés par une modification des activations lexicales des divers candidats en fonction de leur alignement avec des *formes phonologiques* typiques de la langue Une description plus approfondie de cette catégorie de processus est présentée dans la section suivante.

2. Indices de segmentation prélexicaux

Pour se libérer des contraintes liées à des processus de segmentation lexicale qui ne reposeraient que sur des procédures d'accès au lexique, ainsi que pour mieux rendre compte des données comportementales ou acoustiques, un certain nombre de propositions ont été développées dans lesquelles il est envisagé de recourir à des informations prélexicales pour guider la segmentation d'un signal de parole en mots. On trouve, parmi ces propositions, une variété d'indices qui permettraient de localiser des frontières de mots avec plus ou moins de certitude. L'approche la plus courante a consisté à proposer des indices signalant des frontières. Il a également été envisagé, assez rarement il est vrai, de recourir à des indices indiquant la relative nécessité de regrouper certaines séquences à l'intérieur d'une hypothèse lexicale. Ces deux approches fournissent cependant un intérêt équivalent dans la description de modèles de la segmentation lexicale : la première se fixe pour objectif de séparer des séquences en localisant des frontières (ségrégation) alors que la seconde cherche des indices impliquant la nécessité de regrouper certaines portions du signal pour la recherche lexicale (intégration). Quoiqu'il en soit, on peut classer ces indices dans deux catégories : des informations segmentales qui sont fournies par les caractéristiques acoustico-phonétiques ou phonologiques individuelles des sons ; et des informations suprasegmentales qui impliquent la mise en relation d'indices acoustico-phonétiques ou phonologiques portés par plusieurs segments ou groupes de segments de la chaîne parlée. Dans tous les cas, les indices proposés présentent une certaine régularité qui permettrait de les utiliser comme des prédicteurs de la présence ou de l'absence d'une frontière de mots.

2.1. Indices segmentaux

Parmi les indices segmentaux, dont la particularité est de fournir des informations utilisables sans qu'il soit nécessaire de les mettre en rapport avec d'autres éléments du signal, les indices majeurs de frontières lexicales sont les variations allophoniques. Les allophones sont des variantes acoustiques d'un même phonème. Ils n'ont pas de valeur distinctive dans la langue mais présentent des variations acoustiques qui peuvent, dans certaines conditions, adopter un comportement stable. On observe notamment des régularités en fonction de la structure de la syllabe dans laquelle est prononcé le phonème (en français, /o/ s'ouvre et aboutit au phonème [ɔ] dans une syllabe fermée). La prononciation peut aussi dépendre de la position du phonème dans la syllabe (en français, /r/ se prononce différemment en attaque ou en coda). Leur prononciation varie donc en fonction de paramètres précis sans induire de distinctivité phonémique. Ces variantes acoustiques constituent des représentants de la même catégorie phonémique.

2.1.1. En anglais

Les indices qui ont été mis en évidence dans des tâches perceptives se rapportent à la langue anglaise. Nous verrons dans la section suivante que des indices similaires pourraient intervenir en français. Les travaux présentés montrent que certaines régularités allophoniques peuvent être utilisées par des locuteurs de la langue comme indicateurs de frontières de mots. En anglais, il existe par exemple une tendance à prononcer les phonèmes en position initiale de mots avec des caractéristiques particulières qui les distinguent des mêmes phonèmes prononcés en position finale ou médiane. Ces différences de prononciation peuvent être utilisées par des locuteurs pour identifier une séquence de parole ambiguë (Nakatani & Dukes, 1977). Les auteurs montrent, dans une analyse acoustique de séquences lexicalement ambiguës, que le signal de parole est porteur d'indices différenciant les phonèmes selon qu'ils sont prononcés en début de mot ou dans d'autres positions. Ainsi, les consonnes occlusives sont glottalisées ou laryngealisées en position initiale : le /n/ médian de 'no notion' -qui se trouve à l'initiale du second mot- a des caractéristiques acoustiques différentes du /n/ médian de 'known ocean' -qui est localisé à la fin du premier mot. Cette modification des caractéristiques acoustiques des consonnes en fonction de leur position lexicale permet à des auditeurs d'identifier correctement la séquence effectivement produite malgré son ambiguïté phonologique (avec 70 à 90% de taux de correspondance entre la séquence lexicale effectivement produite et la réponse des auditeurs). Par ailleurs, c'est la laryngealisation ou la glottalisation de la consonne en début de mot qui constitue un indice pertinent ; l'absence de cette caractéristique n'induit pas nécessairement que le phonème soit en position médiane ou finale de mot.

Outre les variations allophoniques dépendantes des frontières lexicales, il existe aussi des modifications acoustiques des phonèmes en fonction de leur position dans la syllabe. Church (1987) propose d'intégrer des connaissances sur les régularités des alternances allophoniques liées à la structure syllabique dans les processus de 'segmentation' (plus précisément de *parsing*) lexicale. Ce type d'informations semble effectivement être mis en œuvre par le système de traitement de la parole. Ainsi, Nakatani & Dukes (1977) observent un rôle de la régularité des variations allophoniques du /r/ et du /l/ anglais. Ces deux phonèmes obéissent à des contraintes de prononciation différentes selon qu'ils sont prononcés en attaque ou en coda syllabique et leurs données comportementales montrent que les locuteurs anglophones se fondent sur ce type d'informations pour identifier correctement les séquences ambiguës. Il existe cependant une relation, dans l'échantillon des séquences qu'ils ont analysées, entre les découpages syllabique et lexical. La distribution des allophones de ces deux phonèmes est déterminée par la position dans les syllabes mais, contrairement au phénomène précédent, cette catégorie de variations allophoniques peut en réalité indiquer aussi bien un début qu'une fin de mot puisque découpage syllabique et lexical ne correspondent pas nécessairement. Ce type d'information est donc moins fiable que la laryngealisation des occlusives en début de mot puisqu'il repose sur des contraintes syllabiques. Ces régularités allophoniques ne sont entièrement fiables que si les découpages syllabique et lexical correspondent. Il reste cependant que ce type d'informations peut servir de pondérateur dans les choix effectués par le système de segmentation lexicale et le modèle proposé par Church (1987) implémente un recours à ce type d'informations pour faciliter les processus de reconnaissance des mots.

Il conviendra toutefois de rester prudent quant à l'interprétation que l'on pourrait être tenté de donner de ces données expérimentales. En effet, Nakatani & Dukes (1977) ont bien montré que, placés dans une situation de choix d'une signification lorsqu'ils écoutent des séquences phonologiquement ambiguës, des informations allophoniques peuvent être prises en compte par le système pour faire le choix correct. Ce résultat ne permet cependant pas d'en conclure un quelconque rôle des indices allophoniques dans les processus mis en place en temps réel pour segmenter un signal de parole continue en mots. En effet, lorsque le système cognitif traite un signal de parole dans une situation de communication, les délais dont il dispose pour analyser et comprendre la suite de mots sont très restreints. Les temps de traitement inhérents à ces situations sont extrêmement courts. Or, dans l'expérience de Nakatani & Duke (1977), les participants disposaient de délais de réponse relativement longs et pouvaient donc potentiellement utiliser des indices qu'ils n'auraient pas été en mesure de prendre en compte dans une situation plus contrainte en termes de délais de traitement.

2.1.2. En français

Il existe en français des régularités similaires à celles de la langue anglaise qui incitent à penser que des processus semblables pourraient prendre place dans le système de segmentation de la parole chez les locuteurs de cette langue. Certains phonèmes du français ont par exemple des variantes allophoniques dont les caractéristiques dépendent en partie de leur position dans la syllabe. Le phonème /r/ est la fricative uvulaire voisée [ʁ] en attaque syllabique et la fricative uvulaire non-voisée [χ] en coda. De même qu'en anglais, la position syllabique ne permet pas de prédire avec certitude la position lexicale d'un phonème. Si, dans une séquence constituée de deux mots, le premier se termine par une consonne et le second commence par une voyelle, la consonne finale du premier mot est enchaînée avec la voyelle initiale du second mot et peut subir les règles de syllabation propres à la langue indépendamment du découpage lexical. La position lexicale du phonème ne peut donc pas être totalement déterminée par ce phénomène d'allophonie du /r/ dans le cas d'un signal de parole continue. Celui-ci permettrait cependant, lorsque le système de reconnaissance de la parole rencontre un /ʁ/ voisé de savoir qu'il est en début de syllabe et, par conséquent, qu'il est possible qu'il soit un début de mot ; l'occurrence d'un /χ/ non-voisé indiquant au contraire une fin de syllabe et, potentiellement, une fin de mot. Cette information pourrait être prise en compte, parmi d'autres, pour générer des hypothèses sur le découpage lexical sans que les décisions ultimes reposent uniquement sur cet indice. A notre connaissance, aucune étude n'a cependant été réalisée afin d'étudier avec précision le rôle des indices allophoniques dans les processus de segmentation de la parole en français.

2.2. Indices suprasegmentaux

La majorité des indices qui ont été proposés comme marqueurs potentiels des frontières lexicales (en anglais aussi bien qu'en français ou en néerlandais) sont en fait de nature suprasegmentale. Cette caractéristique implique la mise en relation des informations portées par plusieurs segments ou groupes de segments de la chaîne parlée. La comparaison des indices entre eux, ou les caractéristiques spécifiques d'une *séquence* de segments, fournit une information sur l'éventuelle présence ou absence d'une frontière lexicale. Nous décrirons ici 3 types d'indices suprasegmentaux : métriques, phonotactiques et distributionnels.

2.2.1. Indices métriques

Les indices métriques sont portés par les variations prosodiques du signal de parole. Celles-ci se manifestent par des changements d'intensité, de durée et de hauteur de la fréquence fondamentale (F_0) qui présentent une régularité dans leur mode de succession. Ces alternances métriques dépendent essentiellement de deux paramètres : les informations prosodiques attachées aux représentations lexicales qui déterminent l'accentuation à adopter lorsqu'on prononce un mot et les contraintes prosodiques liées à la structure phonologique de la langue qui ont pour domaine d'application ce qu'on appelle le groupe prosodique -groupe qui n'est pas nécessairement équivalent au mot. Ces deux composantes de la prosodie constituent deux types d'indices suprasegmentaux qui pourraient jouer un rôle dans les processus de segmentation lexicale.

2.2.1.1. Accent lexical

La langue anglaise se caractérise par l'existence de paramètres accentuels liés aux représentations lexicales (Fear, Cutler, & Butterfield, 1995). Dans les mots anglais, on observe une alternance de syllabes accentuées (*strong syllables*, syllabes fortes) et non accentuées (*weak syllables*, syllabes faibles). Cette alternance étant déterminée par les représentations lexicales, elle a un caractère distinctif : la position de l'accent dans une séquence de phonèmes peut déterminer sa signification. Cet accent lexical présente cependant une régularité importante en ce qui concerne la position qu'il occupe dans les mots. Cutler & Carter (1987) montrent ainsi qu'en anglais, 80% des mots d'un lexique informatisé portent l'accent sur leur syllabe initiale. Dans un système de segmentation de la parole en mots qui chercherait des informations acoustiques aptes à indiquer des frontières de mots, la prise en considération de cette régularité permettrait de privilégier les syllabes fortes comme des débuts de mots.

C'est effectivement le type de procédures qui semblent être mises en œuvre par le système de traitement de la parole chez des locuteurs anglais. Cutler & Norris (1988) ont pu mettre en évidence un rôle des syllabes fortes dans les processus de segmentation lexicale avec une tâche de *word-spotting* (extraction de mot). Cette tâche consiste à présenter à des auditeurs des séquences de parole sans signification (des non-mots) dans desquelles il est parfois possible d'isoler une suite de phonèmes désignant un mot dans la langue. La suite /mɑ̃tɛʃ/ par exemple n'a aucun sens en français ; mais si l'on cherche un mot de la langue dans cette séquence, on peut isoler le mot 'menthe' en position initiale. Cutler & Norris (1988) ont introduit ce nouveau paradigme expérimental avec une expérience dans laquelle ils demandaient à des locuteurs anglais de détecter des mots monosyllabiques dans des non-mots bisyllabiques. Les mots étaient tous des monosyllabes accentués (comme la plupart des monosyllabes de classe ouverte, par exemple *mint*, 'menthe') et apparaissaient en position initiale du non-mot. Les caractéristiques accentuelles de la séquence VC qui suivait variaient afin de tester la différence entre l'occurrence d'une syllabe forte (/əɪv/ dans /mintəɪv/) et celle d'une syllabe faible (/əʃ/ dans /mintəʃ/). Les sujets devaient appuyer le plus rapidement possible sur un bouton réponse lorsqu'ils détectaient un mot de la langue dans les non-mots qui leur étaient présentés. Les mots qui apparaissaient au début des non-mots n'étaient pas préalablement indiqués aux participants[15]. Cutler & Norris (1988) montrent que les locuteurs anglais éprouvent plus de difficulté à reconnaître le mot initial lorsque la seconde syllabe est forte (/mintəɪv/) que lorsque celle-ci est

[15] C'est la raison pour laquelle on traduit *word-spotting* par 'extraction' plutôt que par 'détection' de mot.

faible (/mintəʃ/). Il est donc plus difficile d'identifier un mot qui porte l'accent lexical lorsque celui-ci est prononcé dans une séquence *strong-strong* (SS) que dans une suite *strong-weak* (SW). Selon les auteurs, cet effet serait lié à un alignement des activations lexicales sur le début des syllabes fortes. Dans les séquences *strong-strong*, un ensemble de candidats serait activé à partir du /m/, puis un autre à partir du /t/. Du fait de l'alignement d'un second réseaux de candidats sur le /t/ et du recouvrement des deux réseaux, des phénomènes de compétitions lexicales (cf. Section 1.2.2) devraient s'opérer entre les membres de ces réseaux pour aboutir à l'identification du mot *mint*. Ce recouvrement des activations conduirait à éprouver des difficultés pour rattacher le /t/ aux trois premiers phonèmes qui le précèdent (/min/). Au contraire, lorsque la seconde syllabe est faible, seul le phonème initial serait utilisé pour générer les activations et on n'observerait alors aucun conflit entre deux réseaux de candidats partageant un même phonème. Des résultats similaires ont été obtenus avec des locuteurs néerlandais (Vroomen & de Gelder, 1995) effectuant la même tâche.

2.2.1.2. Alternances prosodiques

L'accent lexical ne constitue cependant pas la seule composante prosodique dans le cadre de l'étude des phénomènes intonatifs. Outre les possibilités de stockage lexical des informations prosodiques qui se manifestent en anglais, la chaîne parlée subit des contraintes prosodiques qui sont indépendantes du lexique. Les théories prosodiques développées en phonologie proposent une représentation de la chaîne de parole en termes de composants prosodiques hiérarchisés (Nespor & Vogel, 1983 ; Rossi, 1999 ; Vaissière, 1983) dont le domaine d'application peut s'étendre au-delà des mots individuels. Une formulation typique de la hiérarchie prosodique consiste à regrouper les phonèmes en syllabes, elles-mêmes regroupées en pieds (*feet*). Ces pieds constituent des 'mots phonologiques' (*phonological words*). Ceux-ci ne correspondent pas nécessairement à des 'mots' à l'exemple de ceux que l'on peut trouver dans un dictionnaire mais regroupent le plus souvent des segments provenant de plusieurs unités lexicales. A partir de cette structure prosodique complexe de la phrase, Grosjean & Gee (1987) proposent, dans le même esprit que Cutler & Norris (1988), de faire reposer les processus d'accès lexical en anglais sur les syllabes fortes. Ils envisagent cependant, en raison de l'absence de correspondance biunivoque entre découpages lexical et prosodique, des procédures plus souples dans l'agencement temporel de l'accès aux représentations segmentales intégrées dans les syllabes non-accentuées par le biais de retours en arrière dans les processus d'intégration des syllabes fortes et faibles en mots. Les syllabes accentuées donneraient lieu à un ensemble d'activations lexicales qui ne se restreindraient pas aux syllabes initiales mais à toute syllabe, quelle que soit sa position dans les mots. Cette première étape de génération d'un réseau de candidats donnerait lieu à une étape d'identification des syllabes non-accentuées environnantes -qui pourrait reposer sur des procédures de traitement aussi bien ascendantes que descendantes. La charge de travail induite par les processus de reconnaissance des mots serait restreinte en raison de la limitation des réseaux d'activation aux mots qui contiennent cette syllabe accentuée. Les mots effectivement présents dans l'énoncé seraient alors identifiés par le biais d'une combinaison des informations dérivées des réseaux d'activation et de l'identification des syllabes non-accentuées.

Ce type de procédures pourrait non seulement s'appliquer à l'anglais mais également -en raison de l'universalité des variations intonatives dans les langues- à toute langue du monde, que l'accent ait ou pas de valeur distinctive. En français par exemple, l'accent n'est pas dépendant d'informations stockées dans le lexique mais repose essentiellement sur les contraintes phonologiques mises en œuvre dans la structuration prosodique des énoncés. Son domaine d'application, le groupe prosodique ('syntagme phonologique' ou 'mot phonologique'), le rend dépendant des paramètres intonatifs de la langue sans être lié aux représentations lexicales. Ces variations prosodiques se manifestent par des modifications de hauteur, d'intensité ou de durée relative des syllabes. Dans la phrase, 'Le menuisier a scié une planche et l'a rabotée', on peut distinguer trois groupes prosodiques :

[Le menuisier] [a scié une planche] [et l'a rabotée].

Chacun de ces groupements prosodiques se manifeste par une élévation progressive -suivie d'une diminution- de la F_0 ainsi que par une accentuation de la syllabe finale. Cette accent final se manifeste par un *allongement de la durée* et / ou une *augmentation de l'intensité acoustique* et / ou une *rapide élévation de la F_0* de la syllabe qui le porte. Du fait de ses caractéristiques, l'accent prosodique français semble être déterminé à remplir une fonction totalement différente de celle qu'exerce l'accent lexical anglais. Tout d'abord, il n'est pas porté par des unités initiales mais finales. En outre, son domaine d'application n'est pas le mot mais le groupe prosodique. On pourrait penser que cette dernière caractéristique le rend difficilement apte à fournir des indices pertinents pour la segmentation d'un signal de parole en mots. Il reste cependant un indice potentiel pour la structuration du signal de parole et la reconnaissance des mots si l'on se réfère au modèle proposé par Grosjean & Gee (1987). En effet, il présente une stabilité beaucoup plus marquée que l'accent anglais qui, bien qu'il soit porté dans la majorité des mots par la syllabe initiale, peut être se porter sur d'autres positions lexicales dans un nombre non-négligeable de mots. L'accent français ne présente pas cette instabilité. Toute phrase prononcée par un locuteur natif subit les contraintes intonatives du français et se trouve découpée en groupes prosodiques qu'il est possible de déterminer par une analyse des paramètres acoustiques propres à cet accent (intensité, durées, F_0). L'occurrence et la détection de cet accent indiquerait sans ambiguïté la fin d'un groupe prosodique.

Banel & Bacri (1997) ont utilisé la tâche de *word-spotting* introduite par Cutler & Norris (1988) afin de tester le rôle de ces régularités intonatives dans les processus de segmentation lexicale en français. Les auteurs ont fait écouter à des locuteurs natifs des non-mots de deux syllabes dont certains commençaient par un mot monosyllabique (les autres ne contenant aucun mot de la langue). Les stimuli expérimentaux présentaient 3 types de configuration intonative : *ïambique* (le pattern classique du français, le phénomène d'accentuation étant localisé sur la syllabe finale du stimulus), *trochaïque* (l'inverse du pattern ïambique, la première syllabe portant seule l'accent), et neutre (une *spondée* dans laquelle les deux syllabes du stimulus sont accentuées). Le paramètre intonatif manipulé était la *durée* : les séquences ïambiques étant constituées d'une syllabe courte suivie d'une syllabe longue (pattern court-long), les stimuli trochaïques étant de type long-court et les spondées présentant un pattern long-long. Le pattern trochaïque (long-court) donne lieu à des délais d'extraction du mot initial plus courts que le pattern ïambique (court-long). Dans la suite de phonèmes /lɑ̃pzɔk/, le mot 'lampe' est détecté assez rapidement lorsque la syllabe /lɑ̃p/ est plus longue que la syllabe /zɔk/. Les latences de détection du mot 'lampe' sont au contraire plus longues lorsque la syllabe /lɑ̃p/ est plus courte que la syllabe suivante. Or, du fait de la localisation de l'accent sur la dernière syllabe des stimuli, le pattern ïambique respecte les contraintes intonatives du français. Dans la configuration trochaïque, au contraire, les paramètres intonatifs correspondent à une rupture prosodique entre la première et la seconde syllabe des stimuli. Les auteurs en concluent que les régularités prosodiques du français *exprimées dans les durées syllabiques* jouent un rôle dans les processus de segmentation de la parole en français. La routine de traitement mise en œuvre consisterait à regrouper les segments de la chaîne de parole en fonction des durées relatives des syllabes, les syllabes longues indiquant la fin d'un groupe prosodique permettraient de structurer le signal perçu et d'empêcher toute tentative d'identification d'un candidat lexical qui ne respecterait pas le découpage prosodique induit par l'allongement de la syllabe finale.

2.2.2. Indices probabilistes

Les résultats présentés en faveur d'un rôle des indices probabilistes dans la segmentation du signal de parole en mots ont été obtenus, chez les locuteurs adultes aussi bien que chez des nouveau-nés, avec des expériences d'apprentissage de langues artificielles utilisant divers paradigmes expérimentaux adaptés à l'âge des participants. Chez l'enfant, comme chez l'adulte, on observe une aptitude à traiter les différences de fréquence des séquences de phonèmes qui pourrait être utile dans les processus de localisation des frontières entre mots.

2.2.2.1. Les travaux sur les langues artificielles

L'approche introduite pour étudier le rôle des indices probabilistes dans les processus de segmentation de la parole a consisté à présenter de longues séquences de parole sans signification dans lesquelles on manipule la fréquence d'apparition de certaines suites de phonèmes. Si les probabilités de succession des phonèmes dans le signal sont traitées par le système pour émettre des hypothèses sur les frontières de mots, on s'attend à ce que les participants présentent, dans une tâche ultérieure, des comportements qui reposent sur les différences de probabilité de ces séquences. Des résultats confirmant l'importance des régularités probabilistes ont été mis en évidence chez des nouveau-nés de quelques jours à qui on a fait écouter des séquences de parole sans signification. On observe, dans une procédure de conditionnement induisant les enfants à tourner la tête vers un jouet lorsqu'ils entendent une séquence familière (*head-turning conditioning procedure*), que les stimuli présentant des probabilités transitionnelles élevées donnent lieu -lorsqu'ils sont présentés isolément- à des taux de préférence plus importants que ceux dont les probabilités transitionnelles sont faibles. Des nouveau-nés sont donc en mesure, dès leur premier contact avec le langage, de calculer les fréquences d'apparition de séquences de phonèmes.

Des processus semblables sont observables chez l'adulte. Des locuteurs anglophones ont été exposés pendant quelques minutes à des séquences de parole sans signification dans lesquelles on avait manipulé les probabilités transitionnelles des paires de phonèmes (Saffran, Newport et al., 1996). Dans une phase ultérieure, les participants écoutaient des stimuli extraits des séquences qui leur avaient été présentées. Leur tâche était de donner un 'jugement lexical' consistant à dire s'ils pensaient que le stimulus correspondait ou pas à un mot dans la langue imaginaire qu'ils venaient d'entendre. On observe une tendance à faire reposer les réponses sur les probabilités transitionnelles des suites de deux phonèmes. Les valeurs faibles de probabilités transitionnelles incitent les participants à supposer qu'une frontière de mots sépare les deux phonèmes alors que les valeurs élevées les conduisent à considérer ces suites comme appartenant au même mot. Les auteurs en concluent que des informations de type probabiliste peuvent être utilisées par le système cognitif dans les processus de traitement de la parole. Des propositions similaires se retrouvent chez d'autres auteurs (Brent & Cartwright, 1996; Brent, 1997) qui présentent des simulations mettant en évidence la facilitation des processus de segmentation lexicale liée à l'utilisation de données probabilistes sur les régularités de co-occurrence des phonèmes.

2.2.2.2. Traitement ou mémorisation ?

La phase préalable de traitement d'une longue séquence de parole sans signification dans les études sur le rôle des indices probabilistes dans les processus de segmentation lexicale pose le problème du locus des effets observés. Cette première phase, qui a pour objet de permettre aux participants de calculer la fréquence des groupes de phonèmes, induit en effet des phénomènes d'apprentissage. Il est nécessaire, pour que des effets probabilistes émergent dans la phase test, de mémoriser les probabilités d'apparition des séquences dans la phase d'acquisition. On n'est donc pas en mesure de savoir si, dans le cadre des processus mis en œuvre par le système cognitif dans ces expériences, on met en évidence un recours aux probabilités dans les processus de mémorisation ou de traitement proprement dit. L'interprétation qui a été donnée de ces résultats est que le système cognitif est en mesure d'utiliser les probabilités transitionnelles des suites de deux phonèmes pour localiser les frontières de mots. Ceci impliquerait que les participants utilisent essentiellement l'information apportée par les séquences rares pour repérer les frontières probables. On peut cependant envisager que les effets de facilitation de la reconnaissance dans la seconde phase reposent en réalité sur une mémorisation plus facile des séquences qui sont présentées fréquemment dans la première phase. Les effets probabilistes observés pourraient alors s'expliquer par une mémorisation plus facile de ce que l'on entend souvent plutôt que par la mise en œuvre de processus consistant à localiser des frontières à partir des informations fréquentielles.

Van der Lugt (1999) a étudié l'effet de la fréquence des paires de phonèmes CV et VC dans les processus de segmentation de la parole en utilisant comme indice de probabilité la fréquence de ces suites dans le lexique néerlandais. Du fait de la comparaison de chaînes phonémiques de fréquences diverses dans la langue, il n'est pas nécessaire d'avoir recours à une phase préalable d'apprentissage afin de faire éventuellement émerger des effets probabilistes. C'est une tâche de *word-spotting* qui a été utilisée dans cette expérience. Les participants, des locuteurs néerlandais, devaient identifier des mots d'une syllabe en position initiale de non-mots bisyllabiques. La fréquence des séquences initiales CV et finales VC du mot était manipulée. On s'attend à ce que, si les probabilités d'apparition des paires de phonèmes sont effectivement utilisées par le système de segmentation de la parole, il soit plus facile d'isoler un mot contenant une séquence fréquente qu'un autre contenant une séquence rare. Les données obtenues ne permettent de conclure qu'à un effet limité de la fréquence. Lorsque la fréquence est manipulée dans la syllabe finale (qui n'appartient pas au mot à extraire) ou en position finale du mot (suite VC), on n'observe aucun effet de la fréquence. Un effet de la fréquence émerge cependant en position initiale du mot. C'est donc la fréquence de la séquence initiale CV qui influence les latences d'extraction de mot. Il semble ainsi que les effets probabilistes observés dans les expériences de langue artificielle ne puissent s'expliquer totalement par un rôle de la fréquence dans la segmentation mais soient en réalité interprétables en termes mnémoniques. On remarquera cependant que les chaînes manipulées dans les expériences de Van Der Lugt (1999) présentent des caractéristiques particulières : elles ne se situent jamais à la frontière lexicale. Ainsi, les séquences dont la fréquence est étudiée sont intégrées soit à la partie qui correspond au non-mot, soit à celle qui correspond au mot. Jamais la fréquence des séquences qui chevaucheraient la partie correspondant au mot et le reste du stimulus n'est manipulée alors que c'est spécifiquement la probabilité d'occurrence de cette séquence qui pourrait être le plus utile au déclenchement d'une segmentation fondée sur les probabilités.

2.2.3. *Indices phonotactiques*

Outre les contraintes prosodiques -qui gouvernent les paramètres intonatifs dans la production de la parole- et les régularités probabilistes -qui concernent la fréquence d'apparition d'une chaîne de phonèmes dans la langue-, les langues obéissent à un certain nombre de contraintes dites phonotactiques. Cette classe de contraintes détermine pour chaque langue l'agencement séquentiel des phonèmes dans les syllabes, les morphèmes et les mots. On considérera que les régularités probabilistes dont nous avons parlé dans la section précédente sont distinctes des contraintes phonotactiques dont il est question ici. Rapidement, les régularités probabilistes nous semblent liées à des fréquences d'apparition quelle que soit leur position dans les unités phonologiques classiques (syllabes, morphèmes, mots). Les régularités phonotactiques impliquent au contraire une référence à la position qu'elles occupent dans ces mêmes unités. Une discussion plus approfondie de la notion de contraintes phonotactiques sera développée dans le Chapitre 3 afin de justifier cette distinction. Dans cette section, nous décrivons un certain nombre de données expérimentales qui ont été interprétées comme un reflet du rôle de contraintes phonotactiques ou syllabiques dans les processus de segmentation de la parole et qui pourraient différer de ce qu'on considère comme lié à l'utilisation d'indices probabilistes.

2.2.3.1. Les expériences de détection de syllabe

Mehler, Dommergues, Frauenfelder, & Segui (1981) ont mis en évidence un rôle de la syllabe dans les processus de traitement de la parole avec une tâche de détection de cibles. Celles-ci avaient soit une structure consonne-voyelle CV (par exemple /ba/) soit une structure

CVC (par exemple /bal/). Dans l'une des conditions expérimentales, la cible correspondait à la première syllabe du mot (/bal/ dans 'balcon', /ba/ dans 'balance'). Dans l'autre condition, il n'y avait correspondance qu'entre cette même cible et la séquence de phonèmes initiale (/bal/ dans 'balance', /ba/ dans 'balcon') mais aucune correspondance avec la structure syllabique du mot. Ainsi, le mot 'balcon' peut se découper en deux syllabes : /bal/ et /kɔ̃/. Si la cible indiquée à un participant est la séquence /bal/, il y a correspondance parfaite avec la syllabe initiale du mot. Par contre, si la cible est /ba/, ces deux séquences ne correspondent pas. Quoi qu'il en soit, dans tous les cas la chaîne de phonèmes initiale était appariée avec la cible. Qu'il y ait ou non correspondance syllabique, les participants pouvaient détecter la cible en position initiale du mot. Les auteurs observent des temps de détection plus courts lorsque cible et syllabe initiale correspondent que lorsque seule la séquence phonémique -indépendamment de sa structuration syllabique- est appariée avec la cible. Ce résultat est interprété par les auteurs comme un reflet du statut de la syllabe dans les processus d'accès au lexique : les représentations lexicales seraient stockées sous un format syllabique et le signal de parole devrait être analysé comme une suite de syllabes afin de l'apparier avec des représentations lexicales (Mehler et al., 1981; Segui et al., 1981). C'est également l'une des propositions avancées par Church (1987) qui, à partir d'une réflexion théorique sur le rôle des contraintes allophoniques et phonotactiques, estime que la syllabe serait une représentation intermédiaire adéquate pour faciliter la segmentation lexicale (*parsing*) et l'appariement (*matching*) avec les représentations lexicales.

Plus récemment, il a été montré (Cutler, Mehler, Norris, & Segui, 1986) que ces résultats ne se répliquent pas chez les locuteurs anglais. Dans une tâche identique, les locuteurs anglais présentent des temps de détection similaires qu'il y ait ou non correspondance entre la cible et la syllabe initiale du mot dans lequel elle apparaît. Ces données ont conduit les auteurs à modifier leur interprétation des résultats en insistant sur le rôle de la syllabe dans les processus de segmentation de la parole plutôt que dans les processus de classification du signal de parole préalables à un contact avec les représentations lexicales (Norris & Cutler, 1985; Segui, Dupoux, & Mehler, 1991). De même que les informations prosodiques pourraient être utilisées par le système de traitement de la parole en structurant les percepts linguistiques et / ou en guidant les processus d'activation lexicale, la structure syllabique des chaînes de phonèmes fournirait des indices pour la segmentation lexicale. Ces indices syllabiques seraient utilisés par les locuteurs français, langue dans laquelle la structure syllabique est relativement claire, alors que les locuteurs anglais utiliseraient plutôt des routines de segmentation fondées sur l'accent lexical.

Il a cependant été montré que les effets syllabiques obtenus avec cette tâche ne se répliquent pas lorsque les séquences porteuses sur lesquelles les participants doivent effectuer leurs jugements ne sont pas des mots (Content, Meunier, Frauenfelder, & Kearns, 1996) ; ce qui conduit les auteurs à envisager que la représentation syllabique impliquée dans l'émergence de ces effets ne peut être construite qu'à partir d'une représentation lexicale déjà constituée. On peut également objecter que les données obtenues en faveur d'un rôle de la syllabe dans les processus de segmentation de la parole l'ont été en demandant à des participants de manipuler explicitement des formes linguistiques syllabiques. En effet, la tâche des sujets consistait à donner une réponse concernant l'appariement éventuel entre la cible qui avait été déterminée au préalable et le stimulus qu'ils entendaient. Il est possible, du fait notamment de l'utilisation du concept de 'syllabe' dans l'apprentissage de la lecture mais aussi de la structure même de la langue française dans laquelle le phénomène d'ambisyllabicité est peu courant -alors qu'il est fréquent en anglais- que les locuteurs français éprouvent plus de facilités que les locuteurs anglais dans la manipulation de cibles syllabiques. Les effets observés pourraient alors s'expliquer par des phénomènes post-lexicaux impliquant l'émergence de stratégies de segmentation qui ne seraient pas réellement utilisées dans les processus de traitement de la parole mais se manifesteraient avec une tâche dans laquelle il apparaît que la syllabe est une unité adéquate et relativement maniable sur laquelle fonder les décisions nécessaires à l'accomplissement de l'expérience.

2.2.3.2. Syllabation et détection de phonèmes

Un rôle de la syllabe a cependant été mis en évidence avec un paradigme de détection de phonèmes, ce qui limite les possibilités d'intervention de stratégies de réponse reposant sur une segmentation non-écologique du signal perçu en syllabes. Vroomen & De Gelder (1999) ont présenté à des locuteurs néerlandais des phrases dans lesquelles les participants devaient détecter un phonème déterminé au préalable. Selon les conditions, le phonème-cible pouvait être prononcé en fin de syllabe comme dans la séquence :

'de.boo**t**.die.ge.zon.ke.nis'[16]

dans laquelle la cible est le phonème /t/. Dans cette situation, on observe qu'il y a correspondance entre la frontière syllabique qui sépare /but/ et /di/ et la frontière lexicale qui

[16] Les points introduits dans la représentation orthographique des phrases indiquent les frontières syllabiques.

sépare *boot* (‘bateau’) et *die* (‘qui’). Dans l’autre condition expérimentale, le phonème-cible apparaissait en position d’attaque syllabique comme dans l’énoncé :

‘de.boo.**t**is.ge.zon.ken’

On observe ici une discordance entre la frontière syllabique qui sépare /bu/ de /tiz/ et la frontière lexicale qui sépare *boot* (‘bateau’) et *is* (‘est’). Dans le premier cas, la consonne cible -qui est toujours une occlusive- est suivie d’une consonne ayant le même mode d’articulation (donc une occlusive). Lorsqu’il y a alignement des frontières, cette consonne est au contraire suivie d’une voyelle. Les participants devaient répondre le plus rapidement possible par l’appui sur un bouton réponse lorsque le phonème-cible était effectivement prononcé dans la phrase. Les auteurs observent des latences de détection plus courtes lorsque les frontières syllabique et lexicale sont alignées -c’est à dire lorsque le phonème-cible est en position de coda syllabique- que dans les séquences dans lesquelles il n’y a pas alignement entre ces frontières -quand le phonème à détecter est en attaque. Ils en concluent que c’est la non-correspondance entre découpage syllabique et frontière lexicale qui induit une difficulté à effectuer une segmentation lexicale correcte et nécessite des procédures de resyllabation afin d’être en mesure de localiser correctement cette frontière. Les locuteurs néerlandais auraient donc recours à une segmentation syllabique dans la tâche de détection de phonèmes.

Ce lien entre segmentation syllabique et détection phonémique s’expliquerait par un recours aux représentations lexicales pour extraire le phonème-cible. Il a en effet été mis en évidence que la tâche de détection phonémique peut être réalisée à partir de deux types de traitements. Il est possible de reconnaître un phonème par le biais d’une analyse purement prélexicale du signal : des processus d’appariement entre la représentation auditive du signal acoustique et les représentations phonémiques stockées en mémoire seraient alors mises en œuvre. Une autre procédure peut cependant être développée qui implique un recours aux représentations lexicales. Il est en effet possible de faire reposer la réponse de détection phonémique sur une préalable identification lexicale. L’existence de deux ‘voies’ possibles pour la réalisation de cette tâche a été mise en évidence dans des expériences dans lesquelles on fait varier la position du phonème à détecter et le statut lexical du stimulus qui le porte (Cutler, Butterfield & Williams, 1987 ; Frauenfelder, Segui & Dijkstra, 1990[17]). Dans ces travaux, on observe que les latences de détection de phonème varient à la fois en fonction du statut lexical mais aussi de la position dans le stimulus. Ces deux variables interagissent, l’effet du statut

[17] cf. Section 2.1 du Chapitre 1.

lexical étant plus élevé lorsque le phonème-cible est prononcé en fin de stimulus qu'au début. Ainsi, si la tâche de détection de phonèmes utilisée par Vroomen & De Gelder (1999) peut reposer sur une identification préalable des mots présentés dans la phrase, toute difficulté à reconnaître les mots induira un délai dans les latences des réponses. C'est ce qui semble se passer lorsque les frontières syllabique et lexicale sont discordantes. La non-correspondance entre ces frontières induirait une difficulté transitoire à localiser la frontière lexicale et, par conséquent, à reconnaître le mot qui porte le phonème-cible. Ceci conduirait à accroître le temps nécessaire à la détection du phonème-cible par rapport à une situation dans laquelle les frontières sont alignées. Dans cette dernière condition, la frontière syllabique fournirait un indice de la présence d'une frontière lexicale. Ceci faciliterait la reconnaissance du mot et, de fait, la récupération de l'information phonémique nécessaire à la détection. On voit ici qu'une tâche n'impliquant pas la manipulation de cibles syllabiques permet de mettre en évidence des effets comportementaux interprétables en termes d'un recours à une segmentation syllabique de la chaîne parlée.

2.2.3.3. Régularités phonotactiques et *word-spotting*

McQueen (1998) a étudié le rôle des contraintes phonotactiques dans les processus de segmentation de la parole avec des locuteurs dont la langue maternelle était également le néerlandais. Les contraintes phonotactiques régissent aussi bien l'agencement sériel des sons de parole que les sons qui sont prononçables par un locuteur natif. Celles que manipule McQueen (1998) portent sur les séquences de parole qui sont admises dans la langue en début de mot. Par exemple, en français, /tl/ et /dl/ sont des groupes de consonnes illégaux alors que /tr/ et /dr/ sont dits légaux. Les premiers n'apparaissent pas dans la langue en début de mot alors que les seconds sont attestés dans cette même position. McQueen (1998) distingue les phénomènes de syllabation de ceux liés aux régularités phonotactiques. En effet, les contraintes phonotactiques constituent seulement l'une des sources qui régissent la syllabation d'une séquence de phonèmes. Nous analyserons plus en détails dans le Chapitre 3 la distinction qui peut être faite entre régularités phonotactiques et syllabation. Nous nous contenterons de mentionner ici que les données présentées par Vroomen & De Gelder (1999) avec une tâche de détection de phonèmes et celles obtenues par McQueen (1998) en *word-spotting* peuvent tout à fait refléter l'utilisation d'informations similaires du fait du lien particulier qui existe entre les contraintes phonotactiques d'une langue et les procédures de syllabation des phonèmes dans la chaîne parlée. Par ailleurs, la dernière expérience présentée par McQueen (1998) afin de dissocier indices phonotactiques et syllabiques n'ayant pas donné de résultats concluants, il n'est pas possible de dire si les résultats obtenus sont liés à des contraintes phonotactiques ou à des processus de syllabation de la chaîne parlée.

Afin de montrer que les contraintes phonotactiques constituent une classe d'informations utilisées dans le cadre des processus de segmentation de la parole en mots, McQueen (1998) reprend la tâche de *word-spotting* introduite par Cutler & Norris (1988) en manipulant le statut du groupe de consonnes médian. Les mots à détecter apparaissent soit en position initiale soit en position finale. Dans ces deux conditions, il fait varier la légalité phonotactique du groupe consonantique médian, celui-ci étant soit légal (/vr/, /dr/) soit illégal (*/mr/, */nr/). La tâche des sujets consiste à appuyer le plus rapidement possible sur un bouton réponse lorsqu'ils repèrent un monosyllabe de la langue en position initiale ou finale d'un non-mot de deux syllabes. La comparaison des données en position initiale et finale permet de contrebalancer la relation qui existe dans ces expériences entre légalité phonotactique du groupe de consonnes médian et alignement des frontières phonotactique et lexicale.

Tableau 2 : Statut du groupe de consonnes médian dans l'expérience de McQueen (1998) en fonction de l'alignement entre frontières phonotactique et lexicale.

	Non-alignement	**Alignement**
Position initiale	illégal /pɪlmrem/	légal /pɪlvrem/
Position finale	légal /fidrɒk/	illégal /fimrɒk/

Lorsque le mot à détecter est en position initiale, le groupe légal donne lieu à un non-alignement de la frontière phonotactique avec la frontière lexicale. Si le mot à détecter est *pil* ('pilule'), le stimulus /pɪlvrem/ -qui contient le groupe consonantique médian légal /vr/- donne lieu à une correspondance entre segmentation phonotactique /pil.vrem/ et segmentation lexicale /pil.vrem/. Lorsque le groupe médian est illégal (par exemple */mr/ dans /pɪlmrem/), il n'y a par contre plus correspondance entre segmentations phonotactique /pilm.rem/ et lexicale /pil.mrem/. Au contraire, si le mot à détecter est en position finale on observe une relation inverse entre légalité du groupe médian et alignement des frontières. Pour une détection du mot *rok* ('jupe'), la séquence /fimrɒk/ (dans laquelle */mr/ est illégal) donne lieu à un alignement de ces frontières -avec une segmentation phonotactique /fim.rɒk/- alors qu'un groupe de consonnes illégal (par exemple /dr/ dans /fidrɒk/) induit un non-alignement des découpages phonotactique (/fi.drɒk/) et lexical (/fid.rɒk/). Les informations présentées dans le Tableau 2 résument l'agencement du lien entre légalité et alignement dans cette expérience. Les résultats obtenus lorsque les mots à détecter apparaissent en position initiale montrent que l'absence d'alignement entre les frontières phonotactique et lexicale -qui correspond à l'occurrence d'un groupe illégal en position médiane- donne lieu à des délais d'extraction du mot plus longs que la condition de non-alignement. Le mot *pil* est ainsi détecté plus lentement dans /pɪlmrem/ que dans /pilvrem/. On observe le même effet de l'alignement lorsque les mots à détecter sont en position finale ; la condition de non-alignement correspondant ici à un groupe consonantique médian légal. Dans cette condition, le mot *rok* est détecté plus lentement dans /fidrɒk/ que dans /fimrɒk/. McQueen (1998) en déduit un rôle des indices phonotactiques dans les processus de segmentation lexicale. Ces effets sont simulés (Norris, McQueen, Cutler, & Butterfield, 1997) dans une implémentation du modèle SHORTLIST (Norris, 1994) intégrant des processus de modification des niveaux d'activation en fonction des indices de segmentation phonotactiques.

2.2.3.4. Peut-on en déduire un rôle des régularités phonologiques dans les processus de segmentation lexicale ?

Les expériences réalisées par Vroomen & De Gelder (1999) et par McQueen (1998) aboutissent, avec des tâches qui n'induisent pas la manipulation directe de segments syllabiques de la part des auditeurs, à des résultats concordants en ce qui concerne un potentiel rôle des contraintes phonologiques (qu'elles soient d'origine phonotactique ou syllabique) dans les processus de segmentation de la parole en mots. Selon les auteurs, le système de reconnaissance de la parole utiliserait les connaissances intégrées au cours de l'apprentissage pour émettre des hypothèses sur de probables frontières lexicales ; ce qui faciliterait dans certains cas la reconnaissance des mots. On a vu cependant, dans le Chapitre 1, que des variables interprétées comme mettant en œuvre une catégorie particulière d'information (lexicale, phonologique ou fréquentielle) peuvent être facilement confondues en raison des particularités de la langue et des relations qui peuvent exister entre fréquence des mots, fréquence des diphones utilisés dans ces mots et régularités phonologiques. On peut par exemple envisager que, dans les expériences réalisées par Vroomen & De Gelder (1999) et McQueen (1998), il y ait une corrélation entre les modalités de la variable alignement (qui correspond à un contraste entre groupes de consonnes phonotactiquement légaux et illégaux ou entre séquences occlusive-voyelle et occlusive-occlusive) et la fréquence des séquences de phonèmes qui servent à la comparaison des modalités de cette variable.

Résumé

Nous avons vu, dans ce chapitre, deux approches complémentaires des processus de segmentation du signal de parole en mots. La première est fondée sur des processus d'accès au lexique aboutissant, par le biais de procédures de sélection ou de compétition lexicales, à la localisation de frontières entre les mots de la chaîne parlée. La seconde accentue le rôle d'informations prélexicales comme les régularités phonologiques (liées aussi bien à des indices segmentaux que suprasegmentaux tels que la prosodie et les contraintes phonotactiques) ou distributionnelles dans la mise en place d'hypothèses sur la localisation probable de frontières de mots. Les deux dernières études présentées ont conduit leurs auteurs respectifs à interpréter les résultats obtenus comme des preuves du recours à des connaissances sur les régularités phonotactiques ou syllabiques de la langue dans les processus de segmentation de la parole en mots. Il est cependant difficile de conclure de manière définitive à un rôle des informations phonologiques sans avoir une description adéquate des concepts de contraintes phonotactiques ou de syllabation et de leur expression dans la langue en termes de régularités probabilistes. L'objet du chapitre suivant est de présenter diverses approches des notions de légalité phonotactique et de syllabation tout en abordant la question du lien entre régularités phonologiques et fréquence.

Chapitre 3

Syllabe et légalité phonotactique

SYLLABE ET LEGALITE PHONOTACTIQUE

Chaque langue du monde présente des caractéristiques qui la distinguent des autres langues. Dans le cas de la structure sonore de la langue, objet d'étude de la phonologie, ces spécificités concernent aussi bien l'inventaire des sons individuels que celui des séquences de sons qui sont attestées dans cette langue. Les théories développées en phonologie ont pour objet de décrire les mécanismes qui permettent de transformer une représentation phonologique dite sous-jacente (représentation abstraite de la séquence de phonèmes à prononcer) en une forme de surface (la forme phonétique) correspondant à ce qui est effectivement prononcé. Les modèles proposés définissent donc deux composantes essentielles : la forme des représentations sous-jacentes et les procédures permettant de transformer ces représentations en forme phonétique de surface. Ces règles peuvent être universelles -elles existent dans toutes les langues- ou paramétriques -auquel cas seules certaines langues en sont pourvues. Les représentations phonologiques et les règles qui sont décrites en phonologie peuvent être assimilées à un ensemble de 'connaissances implicites' qui serait stocké par le système cognitif du locuteur (Pierrehumbert, 1990). Cette composante phonologique serait donc partie intégrante d'une base de données -une grammaire- qui serait construite par l'être humain au cours de son développement. Ceci entrouvre la possibilité de proposer une relation étroite entre la grammaire décrite en phonologie et les processus de perception de la parole étudiés par les psychologues cognitivistes.

Nous avons vu dans le chapitre précédent que des données expérimentales obtenues avec les tâches de détection de phonèmes (Vroomen & de Gelder, 1999) et de *word-spotting* (McQueen, 1998) ont été interprétées comme reflétant l'utilisation par le système de traitement de la parole de connaissances sur les régularités phonotactiques ou syllabiques de la langue dans les procédures de segmentation lexicale. Les effets obtenus ont été interprétés différemment par les auteurs respectifs de ces études : Vroomen & De Gelder (1999) invoquent la construction d'une représentation syllabique de l'entrée acoustique qui conduirait à une segmentation du signal de parole ne correspondant pas au découpage lexical lorsque le phonème-cible est en position d'attaque syllabique. Cette discordance dans le découpage induirait un délai dans les latences de détection phonémique. McQueen (1998) fait quant à lui appel à la notion de contraintes phonotactiques pour rendre compte d'un allongement des délais d'extraction de mot lorsqu'il n'y a pas alignement entre les frontières phonotactique et lexicale.

Nous souhaitons d'une part mettre en évidence une certaine similarité des composantes phonologiques qui sont supposées avoir été utilisées par les auditeurs dans ces deux études. Pour ce faire, nous approfondirons la distinction qui a été faite entre contraintes phonotactiques et procédures de syllabation. Cette réflexion sur le lien entre contraintes phonotactiques et syllabation permettra, dans le Chapitre 4, de remettre en question les données comportementales obtenues. Nous commençons par décrire les concepts fondamentaux de la phonologie en présentant diverses approches de la structure des représentations phonologiques et des transformations qui leur sont appliquées. Nous présentons alors l'un des concepts centraux, aussi bien dans le cadre de la phonologie que dans celui des travaux sur la perception de la parole : la syllabe. Cette présentation nous permet de mettre au jour un lien étroit entre la conception hiérarchique propre au concept de syllabe et le caractère linéaire des contraintes phonotactiques qui constituent l'une des sources d'information pour la syllabation des séquences de phonèmes. Du fait de cette relation, les travaux précédemment cités semblent se rejoindre sur le plan des processus cognitifs qui sont à l'œuvre dans l'émergence des effets observés. En outre, le caractère linéaire des contraintes phonotactiques nous conduit à envisager la possibilité que des phénomènes corrélatifs aux contraintes phonologiques -mais indépendants de celles-ci- soient impliqués dans les effets comportementaux mis en évidence. Nous concluons sur la nécessité d'adopter une définition opérationnelle du concept de 'légalité' afin d'aborder plus clairement le problème de l'éventuelle intervention de ce que l'on peut considérer comme des 'connaissances phonologiques' dans les processus de segmentation de la parole en mots.

1. Représentation sous-jacente et forme de surface

La phonologie est une branche de la linguistique qui se donne pour objectif de comprendre la structure sonore des langues du monde. L'un des postulats des travaux réalisés en phonologie est qu'il existe un format de représentation des sons -la forme phonologique- qui diffère de la forme de surface, c'est à dire la manière dont le message est prononcé -la forme phonétique. Le format de représentation préalable à la réalisation effective des sons est appelé la *représentation sous-jacente*. On peut considérer qu'elle constitue une forme abstraite de la séquence de phonèmes qui va, par l'intermédiaire d'un certain nombre de règles appartenant à ce que l'on nomme une *grammaire*[18], se transformer pour aboutir à la forme de surface correspondant aux sons effectivement produits. Classiquement, les théories phonologiques proposent deux classes de modèles dans une formulation de l'organisation des sons des langues : elles définissent d'une part la forme des représentations sous-jacentes, c'est à dire le mode de stockage des mots de la langue dans le système phonologique. D'autre part, elles décrivent les opérations à appliquer sur cette représentation pour en dériver une forme phonétique de surface. On a donc affaire à des modèles qui définissent aussi bien des caractéristiques structurales (la forme des représentations) que fonctionnelles (les opérations à appliquer sur ces représentations). Ces différents modèles de la forme des représentations phonologiques et de sa transformation en une forme phonétique de surface sont essentiels à la compréhension du processus inverse : l'appariement entre une forme phonétique de surface et la représentation phonologique correspondante. Les théories développées reposent pour la plupart sur des analyses de corpora. Depuis quelques années, cette approche a par ailleurs été complétée par des études comportementales inspirées des paradigmes de la psychologie cognitive : jeux de langage (Pierrehumbert & Nair, 1995), détection de phonèmes (Cutler, Butterfield, & Williams, 1987), amorçage (Titone & Connine, 1997) qui sont conduites afin de tester les prédictions des modèles développés en phonologie.

1.1. Représentations sous-jacentes linéaires vs. hiérarchiques

On peut, en ce qui concerne la forme des représentations phonologiques sous-jacentes, décrire deux classes de modèles qui s'opposent sur le niveau de hiérarchisation de ces représentations. Dans le cadre des travaux initiaux de la phonologie générative proposée par Chomsky & Halle (1968), les représentations phonologiques sous-jacentes ont un format

[18] Par analogie avec la grammaire syntaxique.

linéaire. Dans le cadre théorique actuel, on privilégie des représentations multidimensionnelles, non-linéaires.

1.1.1. L'approche initiale de la phonologie générative

Les travaux originaux de la phonologie générative (Chomsky & Halle, 1968) ont tenté de rendre compte du fonctionnement de la phonologie des langues en ayant recours à des représentations sous-jacentes qui, à tous les niveaux de l'analyse phonologique, présentent une simplicité maximale dans leur structure. Cette simplicité concerne aussi bien l'organisation interne des segments (leur description en termes de traits distinctifs) que leur organisation externe en éléments constituants de structures hiérarchiques (syllabe, pied, …) et semble finalement mal adaptée à la description de phénomènes aussi complexes que ceux rencontrés dans les langues humaines. Nous décrivons en premier lieu les choix adoptés pour une représentation linéaire de la structure interne des phonèmes dans ce cadre théorique. Nous passons alors à une présentation du rôle des structures hiérarchiques dans la description de la composante phonologique des langues.

1.1.1.1. Représentation interne des phonèmes

Dans les travaux initiés par Chomsky & Halle (1968) dans *The Sound Pattern of English* (SPE), les segments de la représentation phonologique (les phonèmes) sont décrits comme des séquences de matrices de traits non structurées (cf. Figure 9). Cette absence de structuration induit une indépendance totale entre les traits distinctifs désignant les diverses catégories phonémiques. Le phonème est ainsi envisagé comme une unique matrice de traits, chacune d'elles correspondant à un segment unique de la chaîne de parole (un phonème) et chaque trait pouvant être manipulé indépendamment des autres. Cette conception de la représentation des phonèmes de la chaîne parlée pose des difficultés considérables dans la description des traits

$$
\begin{array}{ccc}
\text{m} & \text{a} & \text{s} \\
\begin{pmatrix} \text{- syllabique} \\ \text{- continu} \\ \text{+ nasal} \\ \text{+ voisé} \\ \dots \end{pmatrix} &
\begin{pmatrix} \text{+ syllabique} \\ \text{+ continu} \\ \text{- nasal} \\ \text{+ voisé} \\ \dots \end{pmatrix} &
\begin{pmatrix} \text{- syllabique} \\ \text{+ continu} \\ \text{- nasal} \\ \text{- voisé} \\ \dots \end{pmatrix}
\end{array}
$$

Figure 9 : Représentation interne des segments dans le cadre de SPE (Chomsky & Halle, 1968).

distinctifs. Ainsi, certaines langues utilisent des sons qui présentent une évolution séquentielle de leurs caractéristiques phonémiques. Par exemple, les affriquées se caractérisent par une articulation initiale occlusive (comme /t/) qui se poursuit par une friction (par exemple /ʃ/) mais correspond à un unique segment dans la langue (le phonème /tʃ/). Les mêmes types de modification séquentielle se retrouvent dans les langues à tons : on observe par exemple des voyelles au sein desquelles le ton est ascendant ou descendant. Un ton ascendant peut être décrit comme la succession d'un ton bas et d'un ton haut. Une théorie linéaire des représentations phonologiques dans laquelle les phonèmes seraient envisagés comme des matrices de traits indépendantes et disposées séquentiellement devrait créer de nouveaux traits *ad hoc* pour chaque forme nouvelle observée. Par exemple, Chomsky & Halle (1968) proposent un trait [± delayed release[19]] pour rendre compte de l'existence des affriquées (qui portent la valeur positive [+ delayed release] alors que les occlusives portent la valeur négative du trait). Il est également possible de créer des traits [± rising tone] ou [± falling tone[20]] pour décrire des tons qui évoluent à l'intérieur d'une même voyelle. Nous verrons plus en détails dans la section suivante les modélisations non-linéaires possibles de cette opération en fonction du cadre théorique adopté.

Outre la nécessité de créer des traits *ad hoc* pour rendre compte de certains phénomènes phonologiques, on observe également que le recours à des représentations sous-jacentes linéaires génère des difficultés spécifiques liées à la possibilité qu'ont les segments de présenter des comportements alternatifs. Ainsi, les phonèmes géminés constituent une classe très particulière de phonèmes qui peuvent se comporter aussi bien comme une séquence de deux segments courts que comme une unique consonne longue (Pulleyblank, 1989). Cette alternance dans le comportement du phonème est difficile à modéliser avec la structure des représentations qui est proposée dans SPE (Chomsky & Halle, 1968). En effet, un phonème géminé doit alors être représenté par un ou plusieurs traits distinctifs qui le caractérisent soit en termes de séquence de segments courts, soit en termes d'un segment long. Il semble délicat de décrire une même géminée avec un ensemble de traits distinctifs différents en fonction de son comportement.

1.1.1.2. Représentation de la séquence de phonèmes

Les tenants du courant théorique SPE ont par ailleurs évité de recourir à des structures hiérarchiques comme la syllabe (cf. Goldsmith, 1990, p.103) ou le pied (cf. Pulleyblank, 1989)

[19] [± relâchement retardé]

pour rendre compte de l'organisation des phonèmes entre eux. On considère ainsi que c'est une chaîne linéaire d'unités qui subit un ensemble de transformations avant de se constituer en une séquence de segments phonétiques prononçables par le locuteur. Par exemple, si l'on reprend l'exemple donné du mot /medəsɛ̃/ (médecin), celui-ci subit un ensemble de transformations qui déterminent sa prononciation. Elles pourraient être les suivantes :

ə → ∅ / VC _ CV (1)
e → ɛ / _ C (2)
[+ voisé] → [- voisé] / _ [- voisé] (3)

Ces trois règles[21] conduisent à dériver de la représentation sous-jacente /medəsɛ̃/ une forme de surface correspondant à la suite phonétique [mɛtsɛ̃]. Il semble intéressant ici de revenir sur la seconde règle présentée pour rendre compte de la transformation de ce mot. Afin d'expliquer le passage d'un /e/ fermé à un [ɛ] ouvert, on doit -si l'on souhaite ne pas avoir recours à la notion de syllabe- définir une règle qui soit en mesure d'induire cette transformation à l'aide d'une information purement séquentielle en rendant compte de toutes les exceptions possibles. La description de cette transformation d'un /e/ fermé en [ɛ] ouvert dans le cadre d'une conception linéaire des représentations phonologiques sous-jacentes se formulerait de la manière suivante dans la règle que nous proposons : un /e/ fermé s'ouvre lorsqu'il précède une consonne. Il est facile de trouver des exceptions à cette formulation. Par exemple, dans le mot /apeti/, la voyelle /e/ est effectivement située entre deux consonnes mais sa forme de surface ne s'ouvre pas. Il est donc nécessaire de modifier la règle (2) afin de préciser les conditions d'application de cette ouverture vocalique. On peut par exemple proposer que la voyelle s'ouvre lorsqu'elle est suivie de deux consonnes :

e → ɛ / _ CC (4)

mais cette règle rencontre encore des exceptions. Le mot /petrir/ se prononce [petrir] alors que le /e/ est suivi de deux consonnes. Il est donc nécessaire d'aller plus loin dans la spécification de la règle en posant que c'est lorsque la voyelle est suivie de deux consonnes obstruantes qu'elle doit s'ouvrir :

[20] [± ton montant] ou [± ton descendant]
[21] Ces règles ne seraient pas nécessairement les mieux adaptées dans le cadre d'un modèle valide de la phonologie du français s'inspirant de la théorie générative. Une quantité considérable de données linguistiques doivent en effet être analysées afin de sélectionner un nombre minimal de règles induisant une quantité minimale d'exceptions.

$$e \rightarrow \varepsilon \,/\, \_\ [+\text{ obstruante}]\ [+\text{ obstruante}] \qquad (5)$$

En réalité, ce processus d'ouverture de la voyelle s'explique très simplement par les caractéristiques de la syllabe qui la porte. En effet, il suffit de dire que la voyelle s'ouvre lorsqu'elle est prononcée dans une syllabe fermée (ou lourde) pour rendre compte de cette transformation. Nous verrons plus en détails les caractéristiques spécifiques de l'entité *syllabe* dans la seconde section de ce chapitre. Néanmoins, avant même d'approfondir la question de ses fonctions et de la description qui en est donnée en fonction des cadres théoriques, on peut déjà définir la syllabe comme une entité typiquement hiérarchique ; selon Goldsmith (1990, p.117) en effet,

> […] syllabation is a process that associates a linear string of segments with a syllable structure[22].

Le terme de 'structure' syllabique désigne l'organisation de la chaîne de phonèmes en une unité d'ordre supérieur. Il semble ainsi nécessaire, afin de prédire l'organisation phonologique du français avec le moins de règles possibles, d'avoir recours à une indication des frontières d'entités hiérarchiques. La nécessité d'utilisation de la notion de syllabe dans le cadre des règles de la phonologie générative implique donc un relatif abandon du refus d'intégrer des notions de hiérarchisation des séquences de phonèmes dans la grammaire phonologique de la langue. Il reste cependant que la description de l'ouverture du /e/ reste parfaitement descriptible, comme nous l'avons vu, avec des règles ne faisant pas intervenir la syllabe.

Clements & Keyser (1983; cité par Pulleyblank, 1989) décrivent cependant un phénomène qui met en évidence une certaine inadéquation des représentations linéaires pour décrire le fonctionnement des langues. En turc, un certain nombre de règles impliquent un environnement phonémique qui peut présenter deux caractéristiques sans rapport l'une avec l'autre : soit la consonne soumise à la règle phonologique doit être en fin de mot, soit elle doit être suivie d'une autre consonne. Prenons par exemple la règle de dégémination des consonnes. Si une consonne géminée précède une voyelle, elle n'est pas modifiée. Si cette consonne est suivie d'une consonne ou si elle est prononcée en fin de mot, elle subit par contre une dégémination. Ces deux conditions constituent, dans le cadre d'une théorie linéaire des représentations phonologiques, deux situations sans rapport l'une avec l'autre. Ces deux environnements, qui ne partagent aucune propriété commune, constituent deux conditions différentes et indépendantes de

[22] La syllabation est un processus d'association d'une chaîne linéaire de segments avec une structure syllabique.

l'application de la règle. Il y a par conséquent disjonction entre deux domaines d'application d'une même règle, l'un phonémique et l'autre lexical. L'existence de disjonctions dans les environnements d'application des règles ne pose pas réellement de problèmes en soi à une théorie linéaire des représentations ; il suffit de créer deux règles distinctes pour rendre compte des deux situations. Cependant, la multitude de situations dans lesquelles Clements & Keyser (1983) observent une disjonction dans l'étude de la phonologie du turc les conduit à remettre en cause la validité d'une approche linéaire.

1.1.2. La phonologie autosegmentale

De nombreux auteurs remettent ainsi en cause une conception purement linéaire des représentations phonologiques sous-jacentes pour proposer des théories non-linéaires regroupées sous le nom générique de *phonologie autosegmentale* (Goldsmith, 1990; Pulleyblank, 1989). Dans la conception initiée par ce cadre théorique, les représentations phonologiques sous-jacentes sont constituées de plusieurs niveaux de représentation qui portent le nom de 'tire' (de l'anglais *tier*)[23]. La ligne de base, le squelette, constitue une suite de positions segmentales sur lesquelles les éléments des autres lignes peuvent s'ancrer. Les autres tires de la représentation (la tire mélodique, la tire syllabique, la tire tonale, …) constituent des niveaux de représentations supplémentaires qui rendent compte des différents niveaux d'analyse et des divers problèmes étudiés dans la structure phonologique des langues. La tire mélodique correspond approximativement à la séquence des valeurs phonémiques stockées à l'intérieur de la représentation. La tire syllabique apporte une information concernant la structuration de cette séquence de phonèmes dans la hiérarchie syllabique. La tire tonale (ou accentuelle) représente les informations intonatives attachées à cette séquence. Les phonèmes sont représentés par des arborescences décrivant leur structure en termes de traits distinctifs qui peuvent présenter des dépendances entre eux. Le lien entre un élément et une position du squelette permet de déterminer l'émergence et les modes de manifestation de cet élément dans la forme de surface. Ces positions squelettales remplissent en effet deux fonctions. Elles permettent aux éléments des autres tires de se manifester lorsqu'ils lui sont liés. Ainsi, seuls les phonèmes ancrés dans le squelette sont exprimés phonétiquement. Elles constituent par ailleurs des unités de temps ou de poids qui, par la structure de leurs connexions avec les autres tires peuvent déterminer par exemple la longueur d'un phonème (Goldsmith, 1990; Pulleyblank, 1989). Cette structuration complexe des représentations phonologiques peut s'appliquer à tout problème de compréhension de la structure sonore des langues. Des problématiques aussi diverses que l'étude des tons des langues africaines (Rialland & Badjimé, 1989) ou la modélisation de la représentation des traits distinctifs (Clements, 1985) peuvent avantageusement utiliser ce type de modélisation pour rendre compte des données linguistiques.

Ce passage d'un cadre théorique linéaire à des modèles pluridimensionnels trouve sa source dans les difficultés rencontrées par les approches initiales de la phonologie générative pour rendre compte des observations linguistiques. On peut dès lors décrire les représentations phonologiques sous-jacentes comme des structures multidimensionnelles dans lesquelles chaque niveau de description phonologique constitue une facette d'un assemblage d'éléments simples en une structure plus ou moins complexe.

1.1.2.1. Représentation interne des phonèmes

Le recours à des représentations sous-jacentes non-linéaires permet notamment de rendre compte de phénomènes tels que le caractère affriqué des phonèmes ou l'existence de tons alternant dans une voyelle unique sans conduire à l'adjonction de nouveaux traits distinctifs au répertoire. C'est la structure des connexions entre les éléments des diverses tires de la représentation phonologique, mieux que l'invention de nouveaux traits *ad hoc*, qui permet de rendre compte de l'existence de ces structures. Ainsi, une consonne affriquée (cf. Figure 10a) est représentée par une consonne unique dans la ligne mélodique (ce qu'elle est effectivement) qui établit des connexions avec les deux valeurs opposées d'un même trait distinctif ([- continu] pour la partie initiale de l'affriquée et [+ continu] pour sa portion finale). De même (cf. Figure 10b), les tons montants ou descendants s'expriment par la connexion d'un ton bas et d'un ton haut sur une position vocalique ; la voyelle portant le ton montant est par conséquent unique (elle occupe une seule position vocalique) mais elle porte les valeurs de deux tons (un ton haut et un ton bas) se succédant pour générer un contour tonal.

En ce qui concerne les phonèmes géminés, ils se représentent tout simplement par une connexion de l'élément mélodique correspondant au phonème avec deux positions squelettales

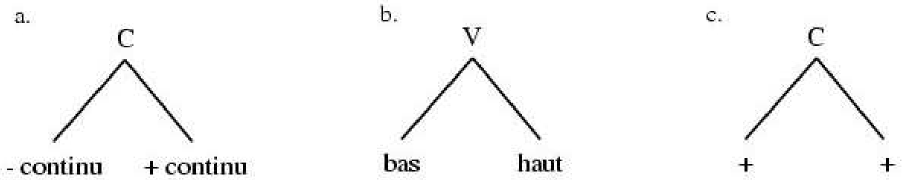

Figure 10 : Représentations non-linéaires décrivant les phénomènes (a.) d'affrication, (b.) de tons montants ou descendants appliqués à un phonème unique ou (c.) de gémination.

[23] On trouve aussi les termes 'ligne' ou 'plan'.

(cf. Figure 10c). Chaque position squelettale représentant une unité segmentale abstraite, il devient inutile d'utiliser des traits spécifiques pour représenter les géminées ; c'est l'ancrage d'un phonème de la ligne mélodique sur deux positions squelettales qui en fait une géminée. Ce mode de représentation facilite par exemple la description du comportement variable des géminées (cf. Section 1.1.1.1 de ce chapitre). En effet, nous avons vu que ces phonèmes peuvent aussi bien adopter le comportement d'une séquence de deux segments courts que celui d'une unique consonne longue. Une représentation interne des phonèmes en termes de matrices de traits nécessite de changer la représentation même du phonème selon l'environnement qui détermine son comportement. Adopter une représentation autosegmentale de la structure des traits distinctifs supprime cette incongruité. Il suffit en effet de dissocier les règles de transformation de la représentation sous-jacente en fonction de la tire sur laquelle elles s'appliquent. Une règle s'appliquant sur la tire mélodique induira un comportement similaire à celui d'une consonne longue alors qu'une règle appliquée sur la tire squelettale déterminera le comportement d'une suite de consonnes courtes. La multidimensionnalité des représentations génère donc une économie considérable dans la description de la structure phonologique des langues. Elle permet en effet de combiner les traits entre eux pour générer des comportements nouveaux. Il est en outre possible d'appliquer les règles de la grammaire sur n'importe quel niveau de la représentation (squelettal, syllabique, tonal, mélodique, …) et de générer ainsi, avec la même règle et la même représentation, des comportements différents.

Il semble important ici de remarquer que cette problématique de la forme des représentations phonologiques est non seulement essentielle dans le cadre de la phonologie -puisque c'est l'une des questions centrales qui doivent être examinées dans toute théorie de la structure phonologique des langues- mais se trouve intrinsèquement liée aux problèmes soulevés par les études de psychologie cognitive portant sur la perception de la parole. En effet, si l'on suppose l'existence d'un lexique mental unique qui pourrait être utilisé aussi bien dans le cadre de la production que de la perception de la parole et si, par ailleurs, on admet que le mode de stockage des mots dans le système cognitif correspond -au moins en partie- au format proposé dans les modèles élaborés en phonologie, les questions posées influencent directement l'angle sous lequel on abordera l'étude des processus perceptifs.

1.1.2.2. Représentation de la séquence de phonèmes

Ce choix d'une structure sous-jacente multidimensionnelle conduit à des descriptions hautement hiérarchisées de la séquence de phonèmes. Il devient en effet possible de structurer la chaîne de phonèmes à l'aide d'arborescences de type syllabique, moraïque, accentuel ou tonal.

Clements & Keyser (1983) montrent par exemple que le choix de représentations sous-jacentes structurées syllabiquement permet de décrire très simplement la possibilité d'observer un même comportement dans des contextes totalement différents. Si l'on revient sur la règle de dégémination présentée dans la section 1.1.1.2, une formalisation linéaire de la représentation phonologique nécessite de recourir à deux règles. La première consiste à dégéminer la consonne lorsqu'elle est suivie d'une frontière lexicale :

$$[+\text{ géminée}] \rightarrow [-\text{ géminée}] \;/\; \_\;\#\# \qquad (6)$$

alors que la seconde effectue la même opération lorsqu'elle est suivie d'une consonne :

$$[+\text{ géminée}] \rightarrow [-\text{ géminée}] \;/\; \_\;C \qquad (7)$$

Dans le modèle proposé par Clements & Keyser (1983), il suffit de recourir à une arborescence syllabique pour rendre compte de cette transformation sans disjonction des domaines d'application. La Figure 11 présente les différents contextes déterminant l'application ou la non application de cette règle de dégémination. La séquence qui subit la transformation est le mot /hakk/ dont la représentation sous-jacente se termine par un /k/ géminé. En (a.), le mot /hakk/ est suivi d'un voyelle ; sa consonne finale conserve donc son caractère géminé. Par contre, en (b.) aussi bien qu'en (c.), cette consonne perd son caractère géminé. Lorsque la consonne est en fin de mot (Figure 11b), c'est à dire lorsque le mot est prononcé isolément, elle est dégéminée en raison de l'impossibilité dans laquelle elle se trouve d'ancrer sa seconde partie dans l'arborescence syllabique[24]. Si elle est suivie d'un mot commençant par une consonne (Figure 11c), elle est confrontée à la même impossibilité et doit également perdre sa gémination. Au contraire, si elle précède une voyelle (Figure 11a), il reste une position d'attaque dans l'arborescence syllabique entre la coda de la première syllabe et le noyau de la seconde syllabe. Il devient alors possible pour cette consonne de s'exprimer phonétiquement puisqu'elle est en mesure de s'ancrer dans la tire syllabique.

Deux ensembles de données favorisent donc le choix de représentations sous-jacentes non-linéaires. Une représentation purement linéaire n'est pas satisfaisante pour décrire la structure interne des phonèmes en termes de traits distinctifs. D'autre part, les règles de transformation de la représentation sous-jacente en une forme phonétique de surface sont grandement simplifiées par la mise en œuvre de structures hiérarchiques comme la syllabe.

[24] Ceci se justifie par le fait qu'une position de coda ne peut porter qu'un seule consonne.

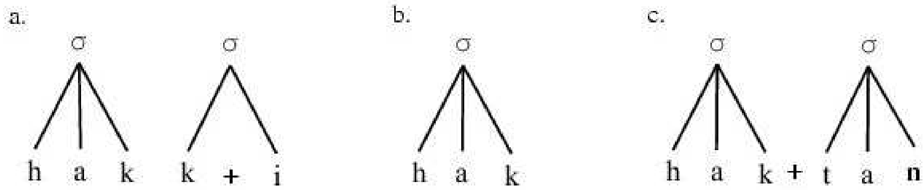

Figure 11 : Représentation de la règle de dégémination du turc dans le cadre de la théorie de Clements & Keyser (1983).

1.2. Processus de choix d'une forme de surface

Outre la modélisation structurale, dont l'objet est de rendre compte de la forme des représentations sous-jacentes, il est nécessaire de décrire le fonctionnement du système qui permet de passer de ces représentations sous-jacentes à une forme phonétique de surface. Or cette conception d'une opération consistant à dériver une représentation d'une autre est actuellement remise en question par les tenants de la Théorie de l'Optimalité.

1.2.1. Transformation d'une représentation sous-jacente

Comme nous l'avons vu depuis le début de ce chapitre, on suppose en phonologie l'existence de deux types de représentations. La première, la représentation sous-jacente, est une entité sur laquelle vont s'appliquer un ensemble de règles de transformation. On peut considérer la somme des représentations sous-jacentes de la langue comme un système qui serait stocké dans le système langagier des locuteurs (par exemple dans le lexique). Le second type de représentations, la forme phonétique de surface, est dérivé des représentations sous-jacentes par l'application des règles de la grammaire phonologique. Cette forme phonétique de surface serait dérivée de ces règles et ne resterait pas stockée par le système phonologique. C'est donc une transformation de la représentation sous-jacente, transformation réalisée par le biais des règles de la phonologie, qui permet d'aboutir à la forme phonétique de surface.

1.2.2. Sélection de la sortie en fonction des caractéristiques souhaitées

Dans le cadre de la Théorie de l'Optimalité, le processus de constitution de la forme phonétique de surface n'est pas une opération de dérivation par règles. Contrairement à la position selon laquelle on transforme une représentation sous-jacente en une forme de surface, la Théorie de l'Optimalité (OT pour Optimality Theory, Prince & Smolensky, 1993), propose que l'on sélectionne la forme de surface parmi l'ensemble des formes de surface possibles. Dans les descriptions de la composante phonologique que nous avons données au préalable, les représentations sous-jacentes étaient soumises à des règles de transformation qui permettaient de parvenir progressivement à une forme correspondant à la sortie phonétique. Dans le cadre de la Théorie de l'Optimalité, on ne parle plus de règles mais de contraintes. Ces contraintes ne servent pas à modifier la structure de la représentation sous-jacente mais à sélectionner le candidat optimal pour la sortie. La composante phonologique ne modifie donc pas ici la représentation phonologique sous-jacente mais choisit un candidat de surface en sélectionnant, parmi l'ensemble des candidats possibles, celui qui respecte au maximum l'ensemble des contraintes définies dans la grammaire. On considère actuellement que les contraintes décrites sont universelles, c'est à dire qu'elles existent dans toutes les langues du monde. Seul leur ordre de priorité permet de différencier les langues entre elles. Ainsi, un changement dans l'ordre des contraintes permet de modifier l'inventaire des formes phonétiques observées dans la langue. Deux classes de contraintes peuvent être proposées : les contraintes *phonotactiques* déterminent une propriété à laquelle doit se conformer la sortie. Dans la seconde catégorie de contraintes, que l'on appelle en anglais *faithfullness constraints*, on considère que la sortie, la forme phonétique de surface, doit être le plus proche possible de l'entrée, la représentation sous-jacente, pour une propriété déterminée.

Après cette rapide introduction aux diverses problématiques rencontrées en phonologie concernant la description interne des segments, leur possible organisation en constituants de structures hiérarchiques et les modalités permettant d'aboutir à une forme phonétique de surface, nous revenons plus longuement sur une structure hiérarchique qui nous intéresse particulièrement dans le cadre de cette thèse : la syllabe.

2. La syllabe

Nous avons entrevu, dans la section précédente, l'intérêt que peut représenter la syllabe pour décrire -dans le cadre d'une théorie générative non-linéaire- la structure des représentations phonologiques sous-jacentes et faciliter la description de règles de transformation de ces représentations en formes de surface. Nous verrons que ce concept est également utilisé dans la Théorie de l'Optimalité. Il reste cependant que, malgré son utilité, la syllabe est un concept mal défini qui, même si l'on y recourt fréquemment, n'en reste pas moins flou (Arnold, 1956). Ce problème s'illustre notamment par l'observation suivante : tout locuteur natif est capable de compter avec précision le nombre de syllabes d'une séquence de parole prononcée dans sa langue… mais il est souvent bien en peine d'indiquer avec certitude la localisation des frontières syllabiques. Ceci semble mettre en évidence l'un des problèmes liés à cette notion de syllabe : on peut facilement pressentir ce concept comme un indicateur de la structuration de la chaîne de parole (Arnold, 1956 utilise l'expression anglaise 'dealing with syllabics' pour désigner le type de connaissances nécessaires à la mise en œuvre de cette capacité) ; mais son analyse en termes de constituants -ainsi que le découpage syllabique lié à l'identification précise des constituants ('dealing with syllables' selon Arnold, 1956)- induisent le recours à des connaissances tout à fait différentes. C'est cette dernière composante, décrire des règles permettant de localiser avec certitude les frontières syllabiques, qui pose le plus problème dans le cadre de la description phonologique de la syllabe.

2.1. Description de la notion de syllabe

La description de l'entité correspondant à ce que l'on appelle la syllabe nécessite d'adopter plusieurs points de vue. En premier lieu, nous verrons à quoi sert la syllabe dans la description de la composante phonologique d'une langue. Nous illustrerons ce rôle par des exemples mettant en évidence la multiplicité des langues dans lesquelles cette notion intervient et la non-adéquation d'une description en termes de séquences purement linéaires de segments. Nous présenterons ensuite diverses conceptions de cette entité. Nous aborderons alors le problème du découpage de la chaîne de segments en syllabes en mettant en évidence les difficultés que pose la localisation des frontières syllabiques. Cette étape de l'analyse nous conduira à poser la question de la validité d'une conception hautement hiérarchique de la structuration syllabique.

2.1.1. A quoi sert la syllabe ?

Cette question nous semble importante dans la mesure où l'approche initiée dans SPE (Chomsky & Halle, 1968) consiste à se dispenser de recourir à des entités hiérarchiques pour décrire la structure phonologique des langues. Il faut donc, avant de poser la question de ce qu'est la syllabe, être raisonnablement certain de son intérêt dans la théorie phonologique. Nous avons vu dans les sections 1.1.1.2 et 1.1.2.2 quelques illustrations mettant en évidence un rôle possible de la structure syllabique dans la description des règles de transformation de la représentation sous-jacente. Dans le premier exemple que nous avons donné concernant l'intérêt de la syllabe pour la description des règles phonologiques, nous avons choisi la règle d'ouverture du /e/ dans une syllabe fermée. Nous avons cependant admis qu'une description adéquate en termes d'une séquence linéaire de segments pouvait parfaitement rendre compte de cette opération. Les travaux de Clements & Keyser (1983) illustrent néanmoins une incongruité relative à la disjonction des environnements d'application des règles du turc dans le cadre d'une théorie qui ferait abstraction de la syllabe.

Si la règle d'ouverture du /e/ en français peut tout à fait s'expliquer à l'aide d'une séquence linéaire de segments et que les règles du turc présentées par Clements & Keyser (1983) peuvent se satisfaire -même si cela semble incongru- d'une disjonction des environnements, un nombre considérable d'autres langues ont été décrites dans lesquelles la syllabe apparaît de nouveau comme une référence essentielle, voire incontournable, de la description phonologique. Dans les langues, la syllabe peut remplir deux fonctions (Goldsmith, 1990). Elle sert d'une part à réguler les règles d'association segmentales, c'est à dire à influencer les conditions d'application des règles de transformation de la représentation sous-jacente. Cette régulation peut s'opérer en fonction de la structure de la syllabe à laquelle appartient le segment qui est soumis à la règle. Ainsi en est-il de la règle d'ouverture du /e/ dont nous avons parlé précédemment pour le français. La même catégorie de phénomènes s'observe en français québécois. Dans cette langue, il existe une règle de relâchement du /i/ dans une syllabe fermée. Par exemple, le /i/ de 'petit' reste tendu dans sa forme masculine ([pti]) mais se relâche dans sa forme féminine qui place le /i/ sous-jacent dans une syllabe fermée ([ptɪt]). La syllabe permet également de réguler les transformations phonologiques en fonction de la place du segment dans la syllabe. En allemand, toute obstruante perd son caractère voisé lorsqu'elle occupe une position de coda syllabique[25]. Alors que la forme féminine *freundin* -'amie'- se prononce [frɔjndən], le /d/ sous-jacent se dévoise dans la forme masculine *freund* -'ami'- dans laquelle il occupe la position syllabique de coda ; *freund* se prononçant [frɔjnt]. Cette catégorie de règles peut difficilement être décrite en faisant abstraction d'une structuration hiérarchique de la séquence de segments.

La syllabe permet d'autre part de déterminer la position lexicale de l'accent. En Yup'ik, une langue parlée dans certaines régions des USA et de l'ancienne URSS, l'accent est porté par la première syllabe fermée. Si la première syllabe d'un mot est ouverte, c'est la syllabe suivante qui porte l'accent. Ce phénomène met en évidence un lien étroit entre les diverses composantes phonologiques des langues. Ainsi, la composante correspondant à la structuration syllabique de la chaîne de segments influence la composante accentuelle. Dans le cadre de la phonologie métrique (*metrical phonology*) par exemple, la syllabe constitue un concept essentiel pour décrire les alternances rythmiques.

[25] coda = fin de syllabe (cf. section 2.1.2.1 pour une description des constituants syllabiques)

2.1.2. *Qu'est-ce que la syllabe ?*

Puisque cette notion de syllabe constitue un élément incontournable de la description des langues, tout recours à ce concept nécessite qu'il soit défini au préalable. Or définir la syllabe est une tâche laborieuse. Le phonème est une notion *relativement* simple à définir dans le cadre de la linguistique étant donné qu'il constitue la plus petite unité distinctive de la langue. Les séquences /ba/ et /ga/ constituant deux formes phonétiques qui s'opposent sur une certaine dimension pour les locuteurs du français, la paire /b/-/g/ présente un statut phonémique en français. Le phonème peut en outre se définir comme le plus petit segment interchangeable dans la langue du fait de l'opposition possible entre /bi/ et /ba/ qui implique par exemple de distinguer les parties initiale et finale de /ba/. Le phonème présente par conséquent une certaine réalité dans la description de la langue puisqu'il en est le fondement même ; décrire une langue consiste à l'analyser en termes de distribution des unités segmentales minimales. La syllabe au contraire n'a pas de statut linguistique clair. On considère classiquement que cette entité constitue une *unité phonologique* à part entière dans le sens où elle joue un rôle dans la description de la langue. Elle ne peut cependant pas être décrite comme une unité segmentale distinctive puisqu'elle en serait elle-même constituée, pas plus qu'il n'est possible d'en donner une description physique en termes de frontières acoustiques ou de pauses articulatoires. Lorsqu'ils tentent de définir le concept de 'syllabe', les auteurs décrivent pour la plupart les règles permettant de découper une chaîne de segments en groupes syllabiques ou présentent leur propre conception de la structure syllabique interne. Par exemple, Hooper (1972) introduit son étude sur le rôle de la syllabe dans la théorie phonologique par ces termes :

> The purpose of this paper is […] to suggest a universal definition [of the syllable] in terms of conventions for the placement of syllable boundaries[26].

Une définition qui nous semble importante est cependant donnée par Fujimura (1990) :

> I think a syllable should be defined as a minimal unit that is utterable in isolation at the phonetic level, and any use of the term should be in some way consistently related to this phonetic notion[27].

---

[26] L'objet de cet article est […] de suggérer une définition universelle [de la syllabe] en termes de conventions concernant la localisation des frontières syllabiques.

[27] Je pense que la syllabe devrait être définie comme l'unité minimale d'articulation au niveau phonétique, et tout usage de ce terme devrait d'une manière ou d'une autre être intimement rattaché à cette notion phonétique.

Nous avons remarqué précédemment qu'aucune pause articulatoire ne permet de localiser des frontières syllabiques dans un signal de parole continue. Il apparaît cependant que, si l'on concentre notre intérêt sur ce qu'est une syllabe prononcée isolément, la conception avancée par Fujimura (1990) fournit des renseignements essentiels. La syllabe constituerait selon cette proposition une unité minimale d'articulation. Cette image nous semble aisément assimilable pour comprendre que cette entité 'syllabe' est centrée sur quelque chose qui peut être articulé en isolation (par exemple une voyelle) et peut se composer, autour ce cet élément articulable, d'éléments qui ne sont pas nécessairement prononçables isolément mais qui, adjoints au précédent, le deviennent. Ceci nous paraît, en élidant la question du statut phonétique ou sous-jacent de l'entité syllabe, un préliminaire correct à une description de la syllabe. Nous n'irons pas plus loin dans cette tentative de définition. Il nous semble plus naturel, du fait de la difficulté intrinsèque à cette tâche, de présenter les diverses propositions qui ont été avancées concernant sa description afin d'appréhender cette notion à partir de la représentation qui en est donnée dans les diverses approches théoriques pour lesquelles c'est l'un des concepts centraux. Nous commençons par présenter deux des conceptions essentielles qui se sont classiquement opposées pour ce qui a trait à ce qu'est une syllabe : l'approche en termes de constituants (*constituent approach*) et celle en termes de courbes (*curve approach*). Nous terminons par la présentation de la théorie moraïque qui, dans une certaine mesure, bouleverse la vision hautement hiérarchisée à laquelle nous sommes parvenus depuis quelques années.

2.1.2.1. Les constituants

On peut décrire la syllabe comme une structure hiérarchique dont les phonèmes sont les *constituants*. Ces éléments simples se regroupent en une unité d'ordre supérieur et y occupent des positions qui possèdent chacune un statut spécifique. Cette spécificité du statut des différentes positions syllabiques conduit à attribuer à chaque élément de l'arborescence une

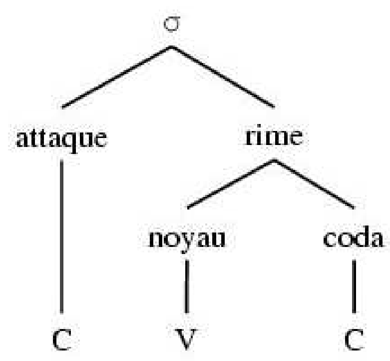

Figure 12 : Représentation hautement hiérarchique de la syllabe en constituants phonémiques remplissant les fonctions d'attaque, noyau et coda.

fonction différente. La représentation syllabique d'un mot adoptera la forme représentée dans la Figure 12. Les constituants syllabiques sont, dans le cadre de la conception la plus répandue de la syllabe (Fudge, 1987) : l'attaque (*onset*, consonne ou groupe de consonnes initial) et la rime (*rhyme*, les phonèmes restants), cette dernière se dissociant en noyau (*nucleus* ou *peak*, généralement la voyelle) et coda (la consonne finale). Cette approche en termes de constituants représente la syllabe comme un groupe de segments phonémiques agencés à l'intérieur d'une structure hautement hiérarchisée et privilégie par conséquent l'aspect interne de cette unité.

2.1.2.2. Les courbes syllabiques

L'une des approches alternatives consiste à définir la syllabe comme une alternance d'accroissement et de diminution d'une variable quelconque de la parole. Par exemple, on peut considérer que la succession des syllabes dans l'activité de parole correspond à une alternance de sons caractérisés par une quantité importante d'énergie (niveau d'intensité acoustique élevé, ce qui est par exemple le cas des voyelles) et de faible énergie (intensité acoustique restreinte comme dans les consonnes). Ces courbes peuvent également se définir en termes articulatoires à partir de la pression engendrée par la constriction (faible pression pour les voyelles et haute pression pour les occlusives). Cette conception de la syllabe se fonde sur une analyse de ses propriétés externes (notamment acoustiques ou articulatoires) plus que sur le statut des éléments la constituant. Elle conduit cependant à l'émergence de difficultés concernant le choix d'un indice acoustique ou articulatoire pertinent (cf. Section 2.2.1.2).

2.1.2.3. La syllabe moraïque

Récemment, un certain nombre d'auteurs (cf. notamment Hyman, 1985; Hayes, 1989) ont proposé, dans le cadre d'une approche en constituants, une conception alternative de la syllabe qui consiste à réduire le niveau de hiérarchisation de l'arborescence tout en conservant une conception non-linéaire de la représentation sous-jacente. On parle, pour désigner cette

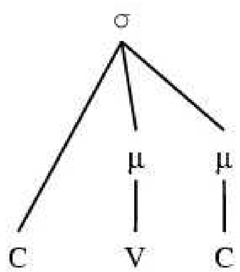

Figure 13 : Représentation moraïque de la syllabe.

proposition, de *structures syllabiques plates*[28] (Pierrehumbert & Nair, 1995). Cette position, connue sous le nom de *théorie moraïque*, consiste à relier directement l'attaque de la syllabe au nœud syllabe. Les autres phonèmes la constituant sont quant à eux liés aux nœuds moraïques : les mores. Il n'existe pas, dans ce modèle de la syllabe, de constituants présentant un niveau de hiérarchisation plus complexe. Dans le cadre de la théorie moraïque, la more constitue une unité de poids. Elle est donc bien plus qu'un constituant se substituant au noyau ou à la coda.

Cette unité permet par exemple de rendre compte très efficacement du comportement des voyelles longues ou de la similarité de comportement entre des syllabes comportant une consonne en position de coda (CVC) et des syllabes portant une voyelle longue (CV:). Dans la Figure 14 on observe que, pour les mêmes mots, les représentations dérivées de la théorie moraïque sont plus simples que celles exprimées en termes de constituants. Ainsi, les deux consonnes initiales du mot français /trak/ n'ont pas besoin de brancher sous le nœud d'attaque et peuvent être directement reliées au nœud syllabique. De même, la représentation d'une voyelle longue comme dans le mot anglais /tri:/ (*tree*, arbre) se passe de branchements doubles (de la voyelle vers le squelette et de ces positions squelettales vers le nœud noyau). Il suffit de faire

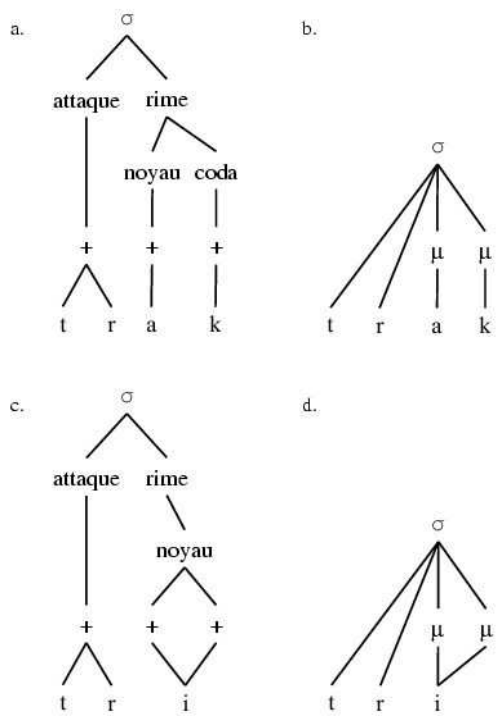

Figure 14 : Comparaison de la représentation de deux mots (l'un français, 'traque' ; l'autre anglais, 'tree') en fonction du choix de la structure syllabique (arborescence hautement hiérarchique en a. et c., syllabe moraïque en b. et d.).

[28] *flatter structures for the syllable*

brancher la voyelle vers les deux positions moraïques (notées μ) pour rendre compte à la fois de son allongement et de sa place dans la syllabe. En effet, une voyelle longue n'occupe pas la position codique mais celle de noyau ; il n'est donc pas envisageable de faire seulement brancher la voyelle vers les positions squelettales et de conserver les connexions du squelette vers les positions de noyau et de coda. Dans la théorie moraïque, poids et position syllabique sont donc intrinsèquement liés. Dans ce cadre là, une syllabe portant une voyelle longue (CV:) se comporte exactement comme une syllabe fermée (CVC). On parle alors, pour mettre en évidence leur similarité, de syllabe 'lourde'. Une syllabe lourde comporte deux mores ; une syllabe légère n'en comporte qu'une. Par ailleurs, le fait qu'une consonne ou un groupe de consonnes en position d'attaque puisse présenter un comportement différent de celui manifesté par la même consonne ou suite de consonnes en coda s'explique par la source du branchement qui provient directement du nœud syllabique en attaque alors qu'il trouve sa source dans les positions moraïques en coda.

2.2. Déterminants de la structuration syllabique

Après avoir présenté les diverses conceptions de la structure interne de la syllabe dans le but de fournir les fondements d'une compréhension de ce que peut être cette entité couramment utilisée en phonologie, il est important de distinguer ce qu'est la syllabe des opérations permettant de localiser ses frontières. Que l'on adopte une approche en constituants aussi bien qu'une approche de courbe syllabique, il faut -afin de segmenter une séquence de phonèmes en syllabes- localiser les frontières qui les séparent. Le processus de syllabation par lequel (1) les segments constitutifs de l'arborescence syllabique acquièrent une position spécifique ou (2) les phonèmes de la chaîne parlée se distribuent en fin ou en début de syllabe en fonction de leurs caractéristiques acoustiques ou articulatoires est donc une composante essentielle du problème qui nous concerne, c'est à dire la capacité qu'auraient les locuteurs natifs de recourir à des connaissances sur la structuration phonologique de la chaîne parlée dans les processus de segmentation de la parole en mots.

2.2.1. Principes d'organisation indépendants des caractéristiques des phonèmes

Plusieurs approches ont été proposées afin de dériver une séquence syllabiquement structurée à partir d'une chaîne linéaire de phonèmes. La première catégorie de propositions adopte l'approche classique des dérivations par règles, celles-ci reposant sur des principes indépendants des caractéristiques intrinsèques des phonèmes (principes de l'attaque obligatoire, de l'attaque maximale). Des approches inspirées de la théorie de l'optimalité envisagent au contraire que le processus de syllabation repose sur la satisfaction de contraintes ordonnées permettant de sélectionner la forme de sortie parmi l'ensemble des candidats possibles. Finalement, un certain nombre d'influences supplémentaires doivent être prises en compte afin de refléter adéquatement les procédures de syllabation de la chaîne de phonèmes : des influences liées à la morphologie, à l'accent (durées vocaliques notamment) et à l'intervention de contraintes phonotactiques.

2.2.1.1. Approche par règles

L'une des approches concernant le mécanisme de structuration syllabique d'une chaîne phonémique consiste à proposer des règles permettant de transformer une forme d'entrée composée d'une chaîne linéaire de segments en une séquence de segments qui acquerraient un statut spécifique dans la hiérarchie syllabique.

2.2.1.1.1. Principe de l'Attaque Obligatoire

Hooper (1972) propose un principe de syllabation selon lequel toute syllabe doit, tant que cette procédure ne viole pas les contraintes phonotactiques de la langue, comporter une attaque. Dans une séquence CVCVC (par exemple /pətit/), ce principe conduit à une syllabation de la suite de phonèmes en CV#CVC car, toute syllabe devant comporter une attaque, la consonne médiane se situe nécessairement à l'attaque de la seconde syllabe. Cette opération ne s'applique cependant pas si la suite médiane CV constitue une séquence phonotactiquement illégale dans la langue. Pour des séquences contenant plus d'une consonne médiane, ce principe n'est pas suffisant. Ainsi dans une suite CVC_1C_2VC, C_2 doit nécessairement appartenir à la seconde syllabe, lui permettant de disposer d'une attaque. Par contre, le statut de C_1 est indéterminé. Elle peut tout aussi bien constituer la coda de la première syllabe que se regrouper avec C_2 à l'attaque de la seconde syllabe. Il est par ailleurs impossible de choisir une solution définitive *a priori* qui consisterait à placer C_1, quelle qu'elle soit, en coda (ou l'inverse) puisqu'en français certaines séquences CVCCV se syllabent en CVC # CV (par exemple, 'taxi', /tak # si/) alors que d'autres présentent une structure CV # CCV (par exemple, 'sacré', /sa # kre/). Des règles supplémentaires sont donc nécessaires pour améliorer la validité de ce principe.

2.2.1.1.2. Principe de l'Attaque Maximale

L'objet du Principe de l'Attaque Maximale (Selkirk, 1982; cf. Goldsmith, 1990 pour une revue) est de fournir les fondements d'une prédiction plus complète des structures syllabiques observées en fonction de la séquence linéaire de phonèmes considérée. Conformément au *Maximum Onset Principle*, l'on doit insérer un nombre maximal de consonnes en position d'attaque syllabique, ceci à condition que cette opération ne viole pas les contraintes phonotactiques de la langue. Dans une séquence CVCCCVC (par exemple, /syrpriz/), il devient possible de décider de la localisation adéquate de la frontière en fonction de la suite de consonnes médiane. L'*Obligatory Onset Principle* ne fournit pas d'information quant au nombre de consonnes à placer en attaque de syllabe, il n'est donc pas possible de choisir entre les diverses configurations possibles. Le *Maximum Onset Principle*, au contraire, contraint un maximum de consonnes à occuper cette position. Ainsi, dans la suite CVCCCVC, les trois consonnes médianes seront positionnées à l'attaque de la seconde syllabe pour autant que les contraintes phonotactiques de la langue le permettent. Si, dans une séquence $CVC_1C_2C_3VC$ provenant d'une langue donnée, les suites C_1C_2, C_2C_3 et C_3V sont phonotactiquement légales, alors la séquence de phonèmes sera découpée en CV # $C_1C_2C_3VC$. Par contre, si C_2C_3 constitue un groupe de consonnes illégal dans cette langue, la suite de segments se découpera en CVC_1C_2 # C_3VC. Le *Maximum Onset Principle* fournit donc une méthode de découpage syllabique plus contrainte que l'*Obligatory Onset Principle*. Elle ne permet cependant pas de prédire intégralement la structuration syllabique de l'ensemble des mots possibles.

2.2.1.2. Optimalité des formes syllabiques

Le développement de la Théorie de l'Optimalité a donné lieu à des propositions spécifiques concernant les procédures de syllabation. Dans ce cadre théorique, on considère que le processus de structuration syllabique correspond à un mécanisme de satisfaction de multiples contraintes qui sont ordonnées selon des critères spécifiques à chaque langue, cette hiérarchie des contraintes permettant de donner lieu à des structures syllabiques différentes en fonction des langues. Tesar (1995) propose un algorithme de syllabation reposant sur 5 contraintes :

ONS	Les syllabes doivent comporter une attaque
NOCODA	Les syllabes ne doivent pas comporter de coda
PARSE	Les segments doivent être rattachés à des positions syllabiques
FILL^{NUC}	Les positions nucléiques doivent être remplies (avec une voyelle)
FILL^{ONS}	Les positions d'attaque doivent être remplies (avec une consonne)

Ces contraintes peuvent être transgressées. La sortie doit seulement être optimale, c'est à dire respecter au maximum l'ensemble de ces contraintes (et notamment celles qui sont le plus haut placées) mais elle peut en transgresser quelques unes. En fonction des langues, leur ordre peut varier, ce qui détermine quelles sont les syllabes optimales dans une langue donnée.

Quel que soit le cadre théorique adopté, on notera qu'une approche cherchant à rendre compte des procédures de syllabation par des méthodes indépendantes des phonèmes en jeu fait des différentes positions syllabiques que sont l'attaque, le noyau et la coda des concepts centraux de son analyse ; induisant de fait une conception hiérarchique de la syllabe. D'autres modèles, dérivés de l'approche en termes de courbes, proposent de faire reposer la procédure de syllabation sur des caractéristiques intrinsèques aux phonèmes. Ces conceptions reposent moins sur le statut hiérarchique de la syllabe que sur sa structuration comme unité phonologique. Elles se situent par ailleurs à l'interface de la phonologie et de la phonétique puisqu'elles insistent sur l'importance des caractéristiques propres des phonèmes dans leur organisation en syllabes.

2.2.2. Principes reposant sur les caractéristiques des phonèmes

Que l'on se situe dans un cadre classique de dérivation par règles (OOP, MOP) ou de satisfaction de contraintes (OT), les mécanismes proposés ne prennent pas en considération les caractéristiques intrinsèques des phonèmes. L'objectif est en effet de prédire la structure syllabique d'une séquence indépendamment de ces caractéristiques. Certaines approches proposent cependant de recourir explicitement à ces caractéristiques afin de prédire la structure syllabique d'une séquence phonémique.

2.2.2.1. Echelle de sonorité

Les deux principes précédents reposent sur des règles d'association syllabique évitant de recourir à des caractéristiques spécifiques des phonèmes considérés. Les procédures de syllabation fondées sur un principe de sonorité (cf. notamment l'approche en termes de courbes) reposent au contraire explicitement sur une description des phonèmes en termes de *sonorité* afin de prédire la structuration syllabique d'une chaîne phonémique (cf. Clements, 1990 ; Klein, 1993). Une syllabe *bien formée* correspondrait à une suite de phonèmes présentant un accroissement puis une diminution *monotones* de la sonorité des segments qui la constituent. On peut définir la sonorité aussi bien selon une approche acoustique qu'articulatoire. Une définition acoustique de la sonorité pourrait ainsi prendre en considération la quantité d'énergie mesurable dans le segment acoustique correspondant approximativement au phonème. Par exemple, une voyelle se caractérise par une enveloppe d'intensité contenant beaucoup d'énergie (cf. Figure 15a) alors qu'une consonne se manifeste par une quantité d'énergie relativement faible (cf. Figure 15b). Des mesures articulatoires pourraient également être mises en œuvre afin d'estimer la sonorité d'un segment : ont par exemple été proposés les indices d'ouverture de l'appareil vocal ou de taille de la chambre de résonance à travers laquelle passe le flux d'air qui donne lieu au segment considéré. Il a également été proposé une échelle de consonanticité qui est

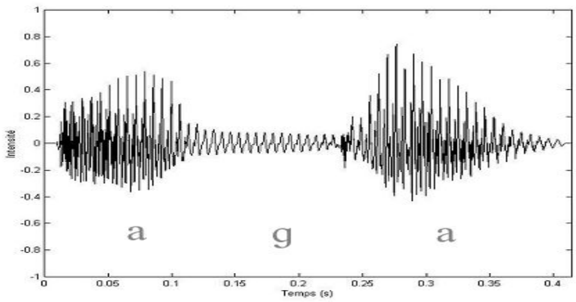

Figure 15 : Enveloppe d'intensité du logatome /aba/ mettant en évidence la différence de quantité d'énergie caractérisant deux classes de phonèmes : les occlusives sonores (comme /g/) et les voyelles.

exactement l'inverse de celle de sonorité (Klein, 1993). Cet indice pourrait quant à lui être dérivé d'une représentation spectrographique du signal reposant sur les caractéristiques du spectre de chaque segment.

A partir d'une comparaison des structures syllabiques observées dans diverses langues, plusieurs échelles de sonorité (ou de consonanticité) ont été proposées qui devraient permettre de prédire la structuration syllabique d'une chaîne linéaire de segments. Un exemple en est donné dans la Figure 16. On observe, de haut en bas, une diminution de la sonorité (ou un accroissement de la consonanticité) en fonction des types de phonèmes. On pourrait, selon cette

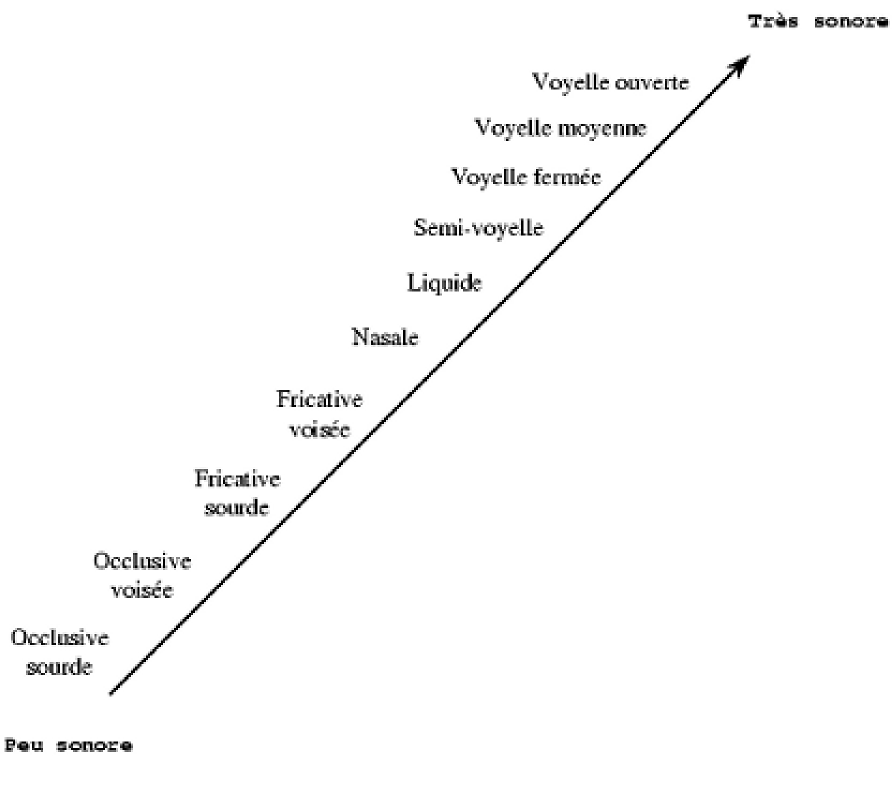

Figure 16 : Distribution de diverses classes de phonèmes selon l'échelle de sonorité.

échelle, déterminer directement à partir des caractéristiques intrinsèques des segments la structure syllabique d'une chaîne de phonèmes. Sur la base d'une échelle donnée de sonorité, la procédure de syllabation consiste à organiser la séquence de segments de telle sorte que le profil de sonorité croisse de façon monotone jusqu'au sommet syllabique puis décroisse alors de manière monotone jusqu'à la coda. Ce principe repose, de même que l'OOP ou le MOP, sur la condition de légalité phonotactique des séquences de phonèmes intégrées dans la syllabe.

En réalité, il apparaît que cette échelle de sonorité ne trouve aucune justification dans les caractéristiques acoustiques ou articulatoires des phonèmes (Clements, 1990). Il ne semble pas possible en effet de déterminer la sonorité d'un phonème à partir d'une analyse de ses caractéristiques physiques. Cette échelle hypothétique n'a été constituée qu'*a posteriori*, elle est déduite de la description des structures syllabiques rencontrées dans plusieurs langues mais, à l'exception de certaines distinctions majeures (par exemple la différence entre occlusive et voyelle), la plupart des classes de phonèmes sont difficilement classifiables à partir de mesures physiques de la sonorité. Par ailleurs, certains phonèmes très semblables sur le plan de leur articulation et de leur niveau dans l'échelle de sonorité présentent des comportements totalement différents dans la structuration syllabique. En français par exemple, les occlusives coronales /t/ et /d/ se comportent tout à fait différemment des autres occlusives bien que l'échelle de sonorité leur attribue un statut équivalent. Par exemple, /t/ ne peut pas être regroupé à l'attaque syllabique avec /l/ alors que l'attaque /gl/ est quant à elle bien formée. Il reste cependant que cette échelle de sonorité permet de modéliser la syllabation d'une chaîne de parole avec plus de précision que les principes de l'Attaque Obligatoire ou Maximale. Clements (1990) en conclut que cette échelle de sonorité, déterminée *a posteriori* et ne correspondant à aucune réalité physique établie, correspondrait non à une variable mesurable avec précision dans le signal (ou au cours de la production) mais à des caractéristiques sous-jacentes des phonèmes, caractéristiques abstraites relevant de la phonologie mais pas de la phonétique. Cette dernière composante fournit cependant les fondements de distinctions primaires entre quelques classes phonémiques (obstruante, liquide et voyelle par exemple).

2.2.2.2. Similarité

Indépendamment des travaux sur la syllabation, Frisch (Frisch, Broe, & Pierrehumbert, soumis; Frisch, 1996) propose d'avoir recours à une mesure de similarité entre phonèmes pour rendre compte des contraintes phonotactiques régulant la co-occurrence des consonnes. Plus deux phonèmes seraient similaires, moins ils auraient tendance à être juxtaposés dans les mots de la langue. Ce taux de similarité pourrait être évalué à partir de la description des phonèmes en termes de traits distinctifs. En outre, les auteurs proposent que cet indice de similarité serait structuré, c'est à dire que certains traits pourraient déterminer le partage d'autres traits. En arabe, langue à morphologie concaténative, la racine verbale est uniquement composée de consonnes. L'obtention d'une forme verbale déterminée passe par l'insertion, au cœur de cette racine verbale consonantique, de voyelles jouant le rôle d'affixe (cf. Figure 17). Une contrainte nommée *OCP-Place* (McCarthy, 1988) empêche de combiner des consonnes homorganiques[29] côte à côte à l'intérieur de la racine verbale, donc de la tire consonantique. Frisch et al. (soumis) observent que cette contrainte n'est pas binaire : elle s'applique avec plus ou moins d'intensité en fonction du taux de similarité des consonnes impliquées et de la distance entre ces consonnes.

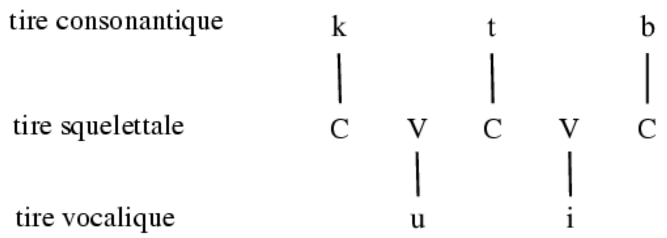

Figure 17 : Représentation du phénomène d'affixation dans la morphologie non-concaténative de l'arabe (extrait de Frisch et al., soumis). La tire consonantique constitue la racine verbale. Les éléments de la tire vocalique sont insérés dans la racine par ancrage à la tire squelettale CV et donnent ainsi lieu à la forme verbale effective.

Les auteurs montrent qu'un modèle assimilant les contraintes phonotactiques de co-occurrence à des contraintes de similarité graduelles rend tout à fait compte de ces données. Partant du principe que la description des phonèmes en termes de traits (acoustiques ou articulatoires) est ancrée non seulement dans la phonologie mais aussi dans la phonétique, cette proposition va donc dans le sens d'un rapprochement de ces deux domaines connexes. Il est par ailleurs tout à fait probable que ce modèle puisse rendre compte des contraintes phonotactiques propres à d'autres langues, par exemple des contraintes de co-occurrence de phonèmes qui seraient sans rapport avec des problèmes de morphologie. Il est en outre possible d'envisager, bien que cette question ne soit pas abordée dans les travaux de Frisch, que cette contrainte de similarité puisse intervenir dans les procédures de syllabation de la chaîne parlée -au même titre que le Principe de sonorité mais à un niveau qui ne serait pas celui de la représentation sous-jacente- tout en reposant sur des caractéristiques effectives des phonèmes mis en jeu dans le processus de syllabation.

[29] Qui ont un ou plusieurs traits en commun.

2.2.3. *Influences supplémentaires sur la syllabation*

On notera qu'un certain nombre de phénomènes conduisent à compliquer la procédure de syllabation qui, finalement et contrairement à ce que l'on peut avoir tendance à penser, ne se résume pas simplement à un processus d'organisation des phonèmes en une structure d'ordre supérieur et qui relèverait uniquement de procédures d'organisation de chaînes phonémiques abstraites. On rappellera en cela la position de Cornulier (1986) qui met en garde contre cette tendance. Parmi ces influences supplémentaires, on compte notamment la structure morphologique des mots et l'accent.

La structure morphologique semble contribuer considérablement au mécanisme de syllabation. Nous avons vu, avec la présentation du *Maximum Onset Principle*, que ce principe contraint un nombre maximal de consonnes intervocaliques à se regrouper à l'attaque de la syllabe suivante. Ainsi, dans une suite CVCCCVC composée de trois consonnes médianes, ce principe devrait conduire, si C_1C_2, C_2C_3 et C_3V sont des séquences phonotactiquement légales, à dériver une structure syllabique CV # $C_1C_2C_3$VC. Il existe cependant en français des mots qui, bien que proches sur le plan des phonèmes qui les constituent, présenteront certainement des structures syllabiques différentes au niveau de leur prononciation. C'est par exemple le cas des mots 'distribuer' ([distribɥe]) et 'déstresser' ([destrese]). Ces deux mots sont constitués, entre leurs deux premières voyelles, d'une suite identique de 3 consonnes : /str/. Or un locuteur natif proposerait certainement de découper 'distribuer' en [dis # tribɥe] avec une frontière syllabique entre /s/ et /tr/, dissociant par conséquent -et en désaccord avec le MOP[30]- les 3 consonnes du groupe médian par le biais d'un isolement du /s/ en fin de première syllabe. Au contraire, et conformément au MOP, le mot 'déstresser' serait probablement découpé en [de # strese], regroupant ainsi l'ensemble des 3 consonnes à l'attaque de la seconde syllabe. La structure morphologique des mots peut donc contribuer notablement à la structuration syllabique d'une chaîne de phonèmes, du moins pour certains groupes de consonnes.

Dans la section 2.1.1 de ce chapitre, nous avons remarqué que la syllabe constituait une composante à part entière des études sur l'intonation et l'accentuation en donnant un exemple du rôle qu'elle pouvait jouer dans le choix du pattern accentuel d'un mot. Il semble cependant que cette relation entre syllabe et accent puisse se définir également dans le sens inverse. Certaines

[30] Cette discordance entre la syllabation et la structure déterminée par le MOP peut sans doute s'expliquer par les caractéristiques du groupe /st/. Il nous semble en effet raisonnable de poser la question du statut effectif de ce groupe de consonnes en termes de 'légalité' (cf. Section 2.2.2.2 pour une discussion de cette notion).

configurations accentuelles peuvent en effet déterminer la structuration syllabique d'une séquence de segments. Par exemple, la durée d'une voyelle contribue au rattachement de la consonne qui lui succède à la syllabe qui contient cette voyelle ou à la suivante. L'évolution de la hauteur fondamentale de la voix est également liée à la structuration syllabique. Grönnum (1999) met en évidence des configurations de F_0 qui dépendent de la structure syllabique des séquences, illustrant ainsi le lien étroit entre structure syllabique et intonation. Treiman & Danis (1988) montrent par exemple qu'en anglais, les consonnes intervocaliques ne sont ambisyllabiques que si la syllabe qui la suit est faible (non-accentuée). Au contraire, si la seconde syllabe d'une séquence bisyllabique et forte (accentuée), la consonne médiane se situe sans ambiguïté à l'attaque de la seconde syllabe.

3. Procédures de syllabation et contraintes phonotactiques

La section précédente nous a permis de passer en revue diverses approches proposées en phonologie pour rendre compte des procédures de structuration syllabique de la chaîne parlée. Notre objectif n'est cependant pas d'étudier la validité de chacune de ces approches mais plutôt d'illustrer en quoi chacune des études comportementales que nous avons citées dans les sections 2.2.3.2 et 2.2.3.3 peut se rattacher préférentiellement à l'une ou l'autre des conceptions théoriques. Nous avons présenté deux approches de la syllabation, chacune constituant à notre avis un point de référence pour les données perceptives illustrant l'interprétation fournie par chacun des auteurs. La première approche envisage la syllabe comme une entité hiérarchiquement structurée qui serait dérivée de règles (ou de contraintes) s'appliquant en partie indépendamment des spécificités acoustiques ou phonétiques des sons impliqués (OOP, MOP, OT). La seconde approche conçoit la syllabe comme une unité qui serait directement déterminée par les caractéristiques propres des phonèmes (principe de sonorité). En dépit des critiques portées à l'encontre des diverses échelles de sonorité proposées et du caractère supposé sous-jacent de ce concept de sonorité, nous avons envisagé que le concept de *similarité* proposé par Frisch (1996) pourrait tout à fait refléter un indice similaire à ceux qui sont envisagés dans les diverses échelles de sonorité (ou de consonanticité) proposés ; indice qui s'ancrerait par contre directement dans une description phonétique et non sous-jacente des segments impliqués.

Un point essentiel et commun à toutes les approches ayant tenté de rendre compte de la syllabation sans faire reposer leurs décisions sur les caractéristiques des phonèmes est qu'en dernière analyse, l'application de ces principes dépend essentiellement de la légalité phonotactique des séquences de segments. En effet, nous avons vu dans la description des principes d'organisation syllabique que sont l'*Obligatory Onset Principle* et le *Maximum Onset Principle* que leur application dépend en grande partie de ces contraintes phonotactiques. Ces principes s'appliquent à condition que les séquences de phonèmes qu'ils manipulent ne soient pas phonotactiquement illégales. Or cette notion de légalité phonotactique, contrairement au concept de syllabe, peut en partie refléter des contraintes linéaires concernant les caractéristiques propres d'une suite de deux segments telles qu'elles sont définies par le principe de sonorité ou par le concept de similarité. Le point que nous souhaitons développer plus particulièrement dans ce chapitre est donc la question du rapport de ces contraintes phonotactiques à la structuration syllabique. Il est par conséquent essentiel de développer une analyse des liens qui existent entre l'entité hiérarchique 'syllabe' et les contraintes linéaires exprimées dans les régularités phonotactiques de la langue. L'objet de cette section est donc de déterminer (1) ce que l'on entend par légalité phonotactique, (2) de comprendre le lien qui existe entre ce phénomène et la procédure de syllabation et (3) de poser la question de la nécessité d'avoir recours au concept de syllabe pour interpréter les données obtenues dans les tâches perceptives. Nous fournissons donc ici une réflexion sur les relations entre syllabe et légalité phonotactique afin de confirmer cette similarité des processus observés dans les effets mis en évidence par McQueen (1998) et Vroomen & De Gelder (1999).

3.1. La notion de légalité phonotactique

Lorsque l'on parle de contraintes phonotactiques, on distingue classiquement deux types de configurations : légales et illégales. Une configuration phonotactiquement illégale consiste en une suite de segments (au minimum deux) qui ne peuvent pas apparaître dans la langue considérée, cette impossibilité se limitant parfois à une position bien précise à l'intérieur des mots ou des syllabes. Par exemple, la séquence /tl/ est dite illégale en français. Elle est *illégale* parce qu'elle n'apparaît pas en début de mot. En effet, le groupe consonantique /tl/ est attesté en français en position médiane de mot (par exemple dans 'atlantique' ou 'atlas'). Par contre, ce groupe n'apparaît jamais en début de mot ; et tout locuteur natif considérerait la suite /tla/ comme *mal formée* en français. On considère en général que cette illégalité provient d'une règle qui interdirait de regrouper deux consonnes coronales à l'intérieur d'une même syllabe. De même, en néerlandais (et probablement aussi en français), /mr/ constitue selon McQueen (1998) une séquence de consonnes phonotactiquement illégale en début de mot. Une séquence de phonèmes illégale serait donc une suite de segments qui ne peut pas apparaître dans certaines positions (ici en début de mot ou de syllabe).

3.1.1. Suites de consonnes : groupes et agrégats

Pulgram (1965) insiste sur la nécessité de distinguer, parmi les séquences de consonnes, ce qu'il nomme 'groupe consonantique' et 'agrégat consonantique'. Cette distinction permet d'opposer les séquences de consonnes en fonction de leur appartenance syllabique. Les 'agrégats consonantiques' constitueraient des séquences tautosyllabiques (qui sont prononcées à l'intérieur d'une même syllabe). Au contraire, les 'groupes consonantiques' constitueraient des séquences de consonnes hétérosyllabiques, chaque segment étant alors rattaché à une syllabe différente. Les *agrégats* seraient donc tautosyllabiques alors que les *groupes* seraient hétérosyllabiques. Cette distinction pourrait s'appliquer aussi bien à des séquences de consonnes différentes (par exemple, /tr/ est tautosyllabique alors que /kt/ est hétérosyllabique) qu'à une même séquence prononcée de deux manières différentes (on se rappellera par exemple les différences de syllabation de la suite /st/ dans 'distribuer' et 'déstresser').

3.1.2. Groupe légal vs. groupe tautosyllabique

La question qui nous importe ici est de déterminer dans quelle mesure ces deux paramètres (légalité et tautosyllabicité vs. illégalité et hétérosyllabicité) sont réellement indépendants l'un de l'autre. Il faut évidemment, pour être en mesure de répondre à cette question, se limiter à des situations dans lesquelles ni la structure morphologique ni les phénomènes accentuels (dont on a vu qu'ils contribuent à la structuration d'une chaîne de segments en syllabes) ne modulent ces phénomènes. Nous souhaitons donc poser la question des relations entre légalité phonotactique et syllabation dans un cas bien précis qui ne peut évidemment pas se généraliser à l'ensemble des situations possibles mais qui suffit à notre avis pour mettre en rapport les données obtenues dans les tâches de segmentation lexicale citées. Il est donc important de signaler ici que la réflexion que nous souhaitons engager est en grande partie éloignée de ce qu'est réellement le processus de génération d'une forme de surface tel qu'il est modélisé en phonétique et phonologie. Nous nous situons d'emblée dans une situation très précise : ce phénomène serait conçu comme relevant essentiellement d'un processus d'association d'une chaîne de segments abstraits avec pour seules influences les mécanismes d'organisation hiérarchique en syllabe et de structuration séquentielle à partir des contraintes de sonorité, de similarité ou des régularités phonotactiques.

3.2. Qu'est-ce qu'une séquence phonotactiquement légale ?

Si le processus de syllabation repose en grande partie sur le respect des contraintes phonotactiques de la langue, il est nécessaire d'être en mesure de déterminer ce qu'est effectivement une séquence phonotactiquement légale. Il est cependant difficile de trouver une définition claire de la légalité phonotactique. En effet, cette notion dépend en grande partie du modèle théorique adopté par les auteurs. Nous présentons ici deux positions différentes qui permettent de définir ce qu'est une séquence légale dans la langue.

3.2.1. Toute séquence de phonèmes attestée à l'attaque syllabique

L'une des définitions possibles de la légalité phonotactique est de considérer comme légale toute séquence de phonèmes qui est possible en début de syllabe dans la langue (Hooper, 1972). Une autre définition possible est de considérer comme légale toute séquence attestée à l'initiale de mot. C'est la définition utilisée par McQueen (1998). Une séquence illégale serait donc constituée de segments qui ne peuvent pas se regrouper à l'initiale de syllabe ou de mot. En réalité, ces deux notions sont intimement liées. C'est sur la mise en évidence de ce lien entre début de mot et début de syllabe que Dell (1995) développe son analyse des groupes de consonnes dans la langue française.

3.2.1.1. Début de syllabe et début de mot

Dell (1995) adopte le raisonnement suivant : toute séquence apparaissant en début de mot constitue également une attaque de syllabe lorsque le mot est prononcé isolément. On peut appliquer le même raisonnement au problème des séquences prononçables en fin de syllabe. La distribution des séquences attestées en début de mot constitue par conséquent un sous-ensemble de celle des séquences attestées en début de syllabe. De même, l'inventaire des séquences pouvant apparaître en position finale de mot constitue un sous-ensemble de celui des séquences apparaissant en fin de syllabe. Dell (1995) choisit donc de recenser les séquences de consonnes qui sont attestées en début et en fin de mot afin de déterminer l'inventaire des groupes de consonnes qui sont en mesure de générer des syllabes 'bien formées'. Selon lui, la distribution des groupes de consonnes constituant des attaques syllabiques bien formées doit permettre de rendre compte des contraintes phonotactiques de la langue (i.e. du caractère légal ou illégal des séquences). Il est évidemment possible que d'autres groupes de consonnes puissent constituer par exemple des attaques de syllabe bien formées mais n'apparaissent pas dans l'inventaire des initiales de mot, cet inventaire constituant seulement une *approximation* des groupes de consonnes qu'il est possible de trouver en attaque ou en coda syllabique. Par conséquent, tout groupe de consonne attesté en début de mot constitue une attaque bien formée alors que l'inverse n'est pas nécessairement vrai : un groupe de consonnes n'apparaissant pas en début de mot ne constitue pas obligatoirement une attaque de syllabe mal formée. L'inventaire des groupes qui ne sont pas attestés en début de mot n'est donc pas assimilable à celui des groupes formant des attaques de syllabe mal formées (i.e. illégales) ; il reste cependant possible -à notre avis- d'estimer l'inventaire des groupes illégaux en considérant que ceux qui ne sont pas observés en début de mot constituent probablement des attaques mal formées.

3.2.1.2. Présentation des données

A partir de cette relation entre la distribution des groupes de consonnes en début de mot et de syllabe, Dell (1995) effectue un comptage des groupes de consonnes attestés en début de mot dans plusieurs corpus de mots isolés de la langue française (notamment Brulex, Content et al., 1990). Selon lui, tout groupe de consonnes attesté en début de mot constitue une attaque de syllabe bien formée. C'est donc l'existence d'un groupe de consonnes en début de mot qui refléterait sa légalité. A partir de l'analyse qu'il fait des différents *corpora* de son étude, il regroupe l'ensemble des groupes de consonnes recensés en 12 classes dépendant de leurs caractéristiques phonétiques mais aussi de ses intuitions concernant la bonne forme (*well-formedness*) ou légalité des séquences. Cette catégorisation permet déjà de voir que Dell (1995) ne conçoit pas le phénomène de légalité comme une distinction binaire entre d'une part des groupes légaux qui pourraient être prononcés à l'attaque d'une syllabe et d'autre part des groupes illégaux ne pouvant pas se regrouper à l'attaque de syllabe. En effet, parmi les séquences attestées en début de mot dans ses corpora, il propose de classer les divers types de groupes de consonnes en deux classes : la classe correspondant aux groupes donnant lieu à des syllabes bien formées (b, c et d) et celle donnant lieu à des attaques de syllabe bien formées *mais* 'déviantes' (e-k). Il existerait donc en réalité 3 classes de groupes consonantiques : les groupes légaux, les groupes illégaux et les groupes légaux mais déviants.

3.2.1.3. La notion de déviance

On peut légitimement se demander ce que signifie cette notion de 'déviance'. Il est clair que la position de Dell (1995) n'est pas de fournir une définition de l'illégalité phonotactique. Il met justement en garde contre cette attitude qui consisterait à conclure, à partir de l'absence d'un groupe dans l'inventaire, à son caractère illégal : il est impossible de savoir si cette absence est un hasard (*accidental gap*) ou est justifiée par les régularités de la langue (*principled gap*). Il reste cependant qu'une position qui consiste à admettre que tout groupe attesté en début de mot constitue une attaque de syllabe bien formée devrait logiquement considérer ces groupes 'déviants' comme bien-formés. Ce qui n'est pas entièrement le cas. En réalité, nous pensons que cette classification dans une catégorie sous le terme de groupes 'déviants' pourrait être liée à la fréquence des groupes de consonnes dans la langue. La notion d'attaque syllabique bien-formée correspondrait alors à quelque chose de plus compliqué qu'un simple recensement des groupes de consonnes attestés en début de mot et nécessiterait une analyse statistique du lexique pour obtenir des données plus fiables.

3.2.2. Toute séquence de phonèmes fréquente

Il serait donc possible, si notre proposition est valide, de définir la légalité d'une séquence en termes de fréquence dans la langue, ce qui impliquerait qu'une séquence peut être plus ou moins légale. C'est la définition qu'en donne Altmann (1990) :

> Phonotactic constraints reflect the different probabilities of co-occurrence of any two or more phonemes.

Il apparaît que les propositions de Frisch et al. (soumis; Frisch, 1996) concernant le rôle de la similarité pour rendre compte de la légalité des suites phonémiques mettent en évidence un lien étroit entre le concept de similarité et la fréquence des séquences dans la langue. Ce lien refléterait une tendance à juxtaposer plus fréquemment des séquences de phonèmes présentant peu de similarité en termes de traits distinctifs. La légalité phonotactique pourrait donc relever d'un phénomène de similarité des phonèmes entre eux, similarité qui se manifesterait par des différences de fréquence d'utilisation de ces séquences dans la langue. Il serait donc possible de déterminer la légalité phonotactique d'une suite de segments à partir de leur taux de similarité ou de la fréquence d'occurrence de cette suite de phonèmes dans la langue.

3.3. Syllabation, contraintes phonotactiques et segmentation lexicale

Les travaux présentés dans le Chapitre 2 (Section 2.2.3) concernant le rôle éventuel des régularités phonologiques dans la segmentation du signal de parole en mots (McQueen, 1998; Vroomen & de Gelder, 1999) fournissent des interprétations alternatives de données comportementales qui sont à notre avis similaires du point de vue des variables manipulées et relèvent de l'opposition entre approche centrée sur la structure hiérarchique de la syllabe et approche centrée sur les caractéristiques des suites de segments. Ces deux interprétations font en effet référence d'une part à un rôle éventuel de la segmentation syllabique (Vroomen & de Gelder, 1999) et, d'autre part, à une segmentation du signal de parole à partir d'indices phonotactiques (McQueen, 1998). Or, comme nous l'avons vu au cours de la description des principes de structuration syllabique (section 2.2), syllabation et contraintes phonotactiques ne sont pas réellement indépendantes l'une de l'autre. L'illégalité phonotactique d'une séquence détermine directement le rattachement des phonèmes qui la constituent à la hiérarchie syllabique. La structure syllabique est donc -en partie- une conséquence des régularités phonotactiques. Les interprétations alternatives proposées par les auteurs de ces études (Vroomen & de Gelder, 1999; McQueen, 1998) sont en fait déterminées par le cadre théorique qui est choisi comme référence. Vroomen & De Gelder (1999) se fondent sur le concept hautement hiérarchique de syllabe pour analyser leurs données alors que McQueen (1998) fait appel à la notion de contraintes phonotactiques pour interpréter les résultats qu'il obtient. On notera que dans une dernière expérience, McQueen (1998) tente de dissocier -sans succès cependant- les rôles respectifs des contraintes phonotactiques et de la syllabation en demandant au locuteur qui enregistre les stimuli de prononcer les séquences médianes, qu'elles soient légales ou pas, de deux manières : soit regroupées à l'attaque syllabique, soit avec une frontière syllabique insérée entre les deux phonèmes. Si aucun effet de la syllabation n'émerge indépendamment de celui de la légalité phonotactique dans la tâche de segmentation lexicale, c'est probablement en raison du lien étroit qui existe entre ces deux phénomènes. Il est probable que la tâche de production d'un groupe illégal à l'intérieur d'une syllabe unique soit particulièrement difficile à accomplir. De même, il est peut-être délicat de contrôler consciemment la prononciation d'un groupe légal dans deux syllabes différentes. Cette difficulté éventuelle de contrôle des phénomènes articulatoires par le locuteur pourrait expliquer l'absence d'effets indépendants et serait déterminée par le lien étroit qui unit contraintes phonotactiques et syllabation.

Nous affirmons donc que les données comportementales obtenues par Vroomen & De Gelder (1999) et par McQueen (1998) sont analysables en termes de mécanismes cognitifs similaires qui reposeraient soit sur une représentation syllabique de la chaîne de phonèmes, soit sur la détection de frontières phonotactiques. Du fait de l'intrication entre régularités phonotactiques et syllabation, il semble difficile de dissocier ces deux composantes du traitement cognitif mis en œuvre par les auditeurs. On pourra objecter que l'expérience réalisée par Vroomen & De Gelder (1999) n'a pas recours à des séquences considérées comme phonotactiquement illégales et que, par conséquent, ces résultats ont nécessairement un lien avec le processus de syllabation. En effet, les auteurs comparent des séquences Consonne-Voyelle dans lesquelles la consonne est nécessairement prononcée à l'attaque syllabique et des séquences Consonne-Consonne qui ne sont pas considérées comme illégales en néerlandais mais induisent néanmoins une syllabation de la première consonne en coda syllabique. Cette proposition n'est cependant valide que si l'on peut affirmer que la légalité est un phénomène binaire, auquel cas l'on est nécessairement amené à mettre en œuvre des procédures spécifiques de syllabation (par exemple le *Maximum Onset Principle*) afin de déterminer la structure syllabique d'un énoncé composé de séquences phonotactiquement légales. En réalité, si l'on se réfère aux travaux de Frisch et al. (soumis), il semble que la légalité d'une chaîne de segments constitue plutôt un continuum sur lequel chaque séquence pourrait se situer. Il serait alors beaucoup plus difficile de dissocier légalité et syllabation, et la distinction de structure syllabique qui est présentée par Vroomen & De Gelder (1999) pourrait être assimilée à une position différente des séquences Consonne-Consonne et Consonne-Voyelle sur un continuum de légalité, les groupes occlusive-fricative étant moins légaux que les suites occlusive-voyelle.

En raison de cette indissociabilité des effets liés à une représentation syllabique de la chaîne de phonèmes ou à la localisation de frontières phonotactiques, nous considérerons désormais que les effets observés dans les expériences présentées (McQueen, 1998; Vroomen & de Gelder, 1999) reposent probablement sur un mécanisme commun qui correspondrait à une segmentation du signal de parole reposant sur des contraintes de nature phonologique (que ces contraintes reposent sur les principes de syllabation ou sur les régularités phonotactiques). Nous choisissons par conséquent d'appréhender la question du rôle des contraintes phonologiques dans les processus cognitifs de segmentation lexicale en nous affranchissant de chercher à distinguer entre structure syllabique et légalité phonotactique. Pour cela, nous considérerons par la suite les diverses conditions expérimentales mises en œuvre dans les expériences comportementales sous l'angle des caractéristiques linéaires des consonnes impliquées plus qu'en fonction de leur tautosyllabicité avérée. Nous adopterons le même principe dans le chapitre suivant lorsque nous présenterons une analyse distributionnelle des groupes de consonnes dans le lexique. Evidemment, nous garderons en permanence à l'esprit que ces caractéristiques phonétiques déterminent en grande partie leur rattachement à la hiérarchie syllabique mais aussi que des groupes présentant des caractéristiques similaires peuvent être syllabés différemment. Cette absence de référence explicite à des structures hiérarchiques nous permettra cependant de proposer des interprétations différentes des effets observés en nous donnant l'occasion de considérer les variables impliquées sous un angle purement linéaire. Les résultats obtenus par McQueen (1998) et par Vroomen & De Gelder (1999) pourraient notamment trouver leur source dans des processus de calcul statistique similaires à ceux qui sont proposés par Saffran, Newport, & Aslin (1996; cf. chapitre suivant pour une justification plus approfondie) plutôt que dans l'intervention de réelles connaissances sur les régularités phonologiques de la langue. Des travaux présentant des simulations connexionnistes du phénomène de syllabation d'une chaîne de phonèmes mettent par ailleurs en évidence la possibilité qu'ont ces modèles de faire émerger des structures syllabiques à partir d'une chaîne linéaire de phonèmes. Ces résultats mettent en évidence la capacité des modèles connexionnistes à développer des propriétés émergentes conformes aux structures syllabiques observées dans les langues aussi bien dans le cadre de la phonologie (Laks, 1995) que de la psycholinguistique développementale (Vroomen, van den

Bosch, & de Gelder, 1998). Typiquement, un modèle connexionniste -ne disposant pas d'informations spécifiques et abstraites concernant ce qu'est une structure syllabique- ne peut que se fonder sur des indices statistiques pour développer des propriétés émergentes ne relevant pas explicitement de principes statistiques (comme la structure syllabique). Il semble donc clair, du fait de cette capacité des modèles connexionnistes à 'apprendre' ce qu'est une syllabe, que des informations probabilistes peuvent être disponibles pour la segmentation et la structuration syllabiques. Il apparaît, à la lumière du lien étroit qui peut exister entre mécanismes de syllabation, contraintes phonotactiques et fréquence d'occurrence des phonèmes impliqués que les interprétations proposées respectivement par McQueen (1998) et Vroomen & De Gelder (1999) pourraient reposer sur des processus cognitifs qui n'auraient pas directement recours à des connaissances phonologiques mais seraient plutôt liés à la fréquence d'occurrence des paires de segments dans la langue. Il est donc nécessaire de conduire une analyse de la fréquence des groupes de consonnes dans le lexique afin de poser les éventuels fondements d'une critique des interprétations proposées par ces auteurs en vérifiant la légitimité du lien qui est envisagé entre régularités phonologiques et fréquence d'occurrence dans la langue.

Résumé

Dans la première partie de ce chapitre, nous avons présenté une introduction aux problématiques rencontrées dans le domaine de la phonologie. Cette introduction nous a permis de mettre en évidence le rôle de représentations hiérarchiques dans les modèles proposés et notamment de celui de 'syllabe'. Nous avons alors entamé une réflexion sur ce qu'est la syllabe et avons insisté sur la distinction qui peut être faite entre sa structure en constituants et certains paramètres linéaires qui pourraient permettre de découper une séquence de phonèmes en syllabes sans avoir nécessairement recours à une conception hautement hiérarchique. L'un des points essentiels qu'il nous a semblé important de soulever est le lien intime qui peut exister entre syllabe et contraintes phonotactiques. Ce concept de contraintes phonotactiques est cependant conçu selon des angles divers par le biais desquels il est possible de concevoir l'opposition entre suite légale et illégale comme une distinction binaire (est légale une séquence attestée dans la langue) ou stochastique (auquel cas il serait possible de définir la légalité comme une échelle continue allant du moins légal au plus légal et qui se manifesterait par la fréquence d'occurrence des séquences dans la langue). Si procédures de syllabation, contraintes phonotactiques et fréquence d'occurrence des segments

sont intimement liées, les effets avancés comme des preuves du recours à des connaissances sur les régularités phonologiques de la langue dans les processus de segmentation lexicale pourraient s'expliquer différemment. Ceci nous conduit à affirmer la nécessité de conduire une analyse distributionnelle des paires de segments dans la langue afin de poser les fondements d'une analyse pertinente des données comportementales présentées.

Chapitre 4

Analyse distributionnelle des séquences de consonnes

Ce chapitre a fait l'objet de deux communications orales dont une dans un congrès avec comité de lecture assorti d'une publication dans des actes :

XXII^{èmes} Journées d'Etude sur la Parole, 15-19 Juin 1998, Martigny, Suisse.

Journée 'Langage et Lexique', 6 Novembre 1999, Institut des Sciences Cognitives, Lyon, France.

ANALYSE DISTRIBUTIONNELLE DES SEQUENCES DE CONSONNES

Le chapitre précédent nous a conduit à mettre en évidence deux approches de la syllabation : la première consiste à considérer la syllabe comme une structure hiérarchique dont la localisation des frontières reposerait essentiellement sur des paramètres indépendants des caractéristiques des phonèmes impliqués (OOP, MOP, OT). Cette approche attribue une importance considérable aux fonctions d'attaque, de noyau et de coda syllabiques que peuvent prendre les phonèmes. Le second type de propositions repose sur les caractéristiques propres des phonèmes en jeu (principe de sonorité notamment) ; le concept de similarité invoqué pour rendre compte des contraintes phonotactiques dans la racine verbale de l'arabe pourrait aussi -à notre avis- constituer un modèle de la syllabation en envisageant que les séquences de deux phonèmes tendent à être le moins similaires possibles à l'intérieur d'une syllabe ; deux consonnes partageant des traits communs tendraient alors à être prononcées dans deux syllabes différentes. Parmi les approches de la première catégorie, la plupart ne s'appliquent que si les contraintes phonotactiques de la langue ne sont pas transgressées. Il existe donc un lien étroit entre contraintes phonotactiques et syllabation. Nous ne disposons cependant pas d'une définition précise et communément admise de la légalité phonotactique. Deux catégories de travaux nous ont conduit à opposer des conceptions alternatives de cette notion : certains auteurs adoptent une conception 'binaire' de la légalité, toute séquence de phonèmes étant légale ou pas. Dans le cadre de cette conception, il est admis qu'une séquence constitue une suite légale si elle est attestée dans la langue en début de syllabe. Nous avons mentionné le problème que posent certains groupes de consonnes cités par Dell (1995) comme *légaux mais déviants* malgré leur existence en début de mot (donc en début de syllabe). Le second type de propositions concernant la légalité phonotactique d'une séquence consiste à l'envisager en termes de *similarité structurée* des segments entre eux. Plus deux segments seraient similaires, moins ils auraient tendance à être juxtaposés à l'intérieur d'une même syllabe en raison de leur plus grande illégalité phonotactique. Or cette similarité des segments semble être inversement corrélée avec leur fréquence d'utilisation dans la langue. C'est notamment le cas dans la morphologie de l'arabe. Si ce lien peut être mis en évidence dans le cadre de la syllabation des phonèmes, on peut s'attendre à ce que les séquences très peu similaires (consonne-voyelle par exemple) soient beaucoup utilisées alors que les séquences les plus similaires (une suite de deux occlusives par exemple) seraient peu utilisées. Cette dissociation entre deux conceptions de la légalité ainsi qu'un lien étroit entre légalité phonotactique et syllabation nous conduit à poser la question d'une définition opérationnelle de la légalité des groupes de consonnes en attaque syllabique, c'est à dire de leur taux de bonne formation (*well-formedness*), en envisageant deux alternatives : on peut considérer d'une part comme légal *tout groupe attesté dans la langue en attaque syllabique*. Selon la démonstration effectuée par Dell (1995), il est possible d'assimiler l'ensemble des groupes de consonnes attestés en début de mot à un sous ensemble des groupes de consonnes qui constituent une attaque de syllabe bien formée. La distribution des groupes de consonnes attestés en début de mot nous permettrait donc d'estimer celle des groupes constituant des attaques syllabiques bien formées. Le modèle proposé par Frisch et al. (soumis) consiste quant à lui à définir la légalité phonotactique de suites de deux phonèmes en termes de similarité. Cette similarité s'exprimerait en outre par des différences de fréquence d'apparition des séquences dans la langue. D'autres travaux (Laks, 1995; Vroomen et al., 1998) montrent par ailleurs qu'une structuration syllabique de séquences linéaires de phonèmes peut être développée par des modèles connexionnistes. Partant de l'observation que ce type de modèle se fonde nécessairement sur des informations probabilistes disponibles au cours de l'apprentissage, il est possible d'affirmer que cette structuration syllabique n'est en mesure d'émerger que grâce à l'utilisation de processus et d'informations purement stochastiques. Si l'on s'affranchit de contraintes morphologiques et accentuelles qui influencent fortement la syllabation, il est alors envisageable de considérer que contraintes syllabiques et phonotactiques s'expriment de façon similaire par la fréquence d'occurrence des suites de segments dans la langue.

Nous avons donc décidé de conduire une analyse distributionnelle des séquences de phonèmes dans la langue et plus particulièrement des groupes de deux consonnes afin d'évaluer la validité de ces interprétations. Si l'on est en mesure de mettre en évidence un lien entre fréquence et syllabation (ou légalité phonotactique), nous serons conduit à mener une réflexion plus approfondie sur la contribution des régularités phonologiques dans les processus de segmentation de la parole (McQueen, 1998 ; Vroomen & de Gelder, 1999). En effet, si la structure phonologique des séquences de consonnes est liée à la fréquence d'occurrence dans la langue, il sera nécessaire de préciser quels processus sont à l'œuvre dans l'émergence des effets observés.

1. Présentation du corpus

Le corpus utilisé pour l'analyse distributionnelle des séquences de consonnes en français est une base de données lexicale informatisée couramment utilisée en psycholinguistique : BRULEX (Content et al., 1990). Nous avons aussi utilisé ce corpus pour les diverses recherches lexicales effectuées lors de la constitution du matériel expérimental de ce travail de thèse. C'est également l'un des corpora analysé par Dell (1995) pour son recensement des groupes de consonnes attestés dans la langue française.

1.1. Caractéristiques de la base de données et méthode d'analyse

BRULEX (Content et al., 1990) est une base de données informatique de mots isolés. Chaque entrée correspond à un mot. Cette base de données recense notamment, pour 35746 mots de la langue française, leur orthographe, leur transcription phonémique et pour la plupart d'entre eux (qui sont au nombre de 26413) leur fréquence d'apparition dans un corpus écrit. Il est ainsi possible, avec cet outil, de faire des recherches lexicales à partir de critères multiples au nombre desquels les structures orthographique et phonémique ou la fréquence des mots mais également de dériver -avec les algorithmes adéquats- d'autres catégories d'informations comme le nombre de compétiteurs des mots, la fréquence des diphones qui les constituent, etc. Nos analyses ont été dans tous les cas effectuées sur les champs de la base de données correspondant à la transcription phonémique des mots.

L'ensemble des analyses présentées ici a été réalisé à l'aide du langage de programmation AWK[31] (Aho, Kernighan, & Weinberger, 1988). Ce langage permet de consulter très simplement des fichiers au format texte et de faire toutes sortes d'opérations sur ces fichiers. La version de BRULEX (Content et al., 1990) que nous avons utilisée est donc une version texte de la base de données originale (qui est fournie dans un format binaire propriétaire).

1.2. Problèmes posés par le choix de ce corpus

Le choix du corpus BRULEX (Content et al., 1990) entraîne un certain nombre de difficultés pour l'interprétation ultérieure des données obtenues. Deux problèmes nous semblent essentiels dans le cadre d'une analyse de la fréquence des groupes de consonnes. Le premier a trait au type de transcription fourni dans la base de données alors que le second est lié au choix d'une base de données lexicale ne permettant pas, par définition, de réaliser un comptage de tous les groupes de consonnes qu'il est possible de rencontrer dans les énoncés linguistiques produits par les locuteurs, c'est à dire dans des phrases.

1.2.1. Transcription des sons

La transcription des séquences de sons de chaque mot qui est fournie dans BRULEX (Content et al., 1990) peut poser des problèmes pour l'interprétation des résultats obtenus. En effet, on peut dire qu'elle ne correspond ni à une transcription phonologique ou phonémique (correspondant à ce que l'on a appelé dans le Chapitre 3 la représentation sous-jacente) ni à une transcription phonétique qui pourrait refléter la manière dont le mot est effectivement produit par un locuteur natif. Par exemple, cette transcription distingue les /o/ ouvert ([ɔ]) et fermé ([o]). Cette opposition n'est pas distinctive en français ; les formes [ɔ] et [o] constituent en effet des allophones du même phonème. Le choix de retranscrire le /o/ de cette manière correspondrait donc à une transcription de sa forme phonétique de surface. De même, la règle de chute du schwa est appliquée pour la transcription. Ainsi, le mot 'paquetage' est retranscrit /paktaʒ/, ce qui implique que le schwa sous-jacent ait été supprimé. Ce paramètre est particulièrement important pour l'analyse des groupes de consonnes attestés dans la langue puisqu'il permet d'observer l'occurrence de groupes de consonnes qui ne seraient pas détectés avec une analyse de la forme orthographique ou phonologique des mots. Par contre, la règle de dévoisement des

[31] Le langage AWK est un langage interprété. Les scripts écrits avec ce langage (cf. Annexes pour des exemples) sont 'interprétés' par un logiciel dédié. Celui-ci existe pour diverses catégories d'ordinateurs et de systèmes ➥

occlusives lorsqu'elles précèdent une constrictive non-voisée n'est pas toujours appliquée. Ainsi, bien que 'médecin' subisse la chute de schwa entre /d/ et /s/, le /d/ n'est pas retranscrit en [t], ce qui donne la transcription /mɛdsɛ̃/ et non pas /mɛtsɛ̃/ comme on aurait pu s'y attendre (cf. Chapitre 3, section 1.1.1.2). Ce choix correspond, contrairement à celui effectué pour la transcription du /o/ et la prise en compte de la chute de schwa, à une représentation plutôt sous-jacente du phonème correspondant à la lettre 'd' dans le mot 'médecin'. Par contre, le mot 'absolu' est retranscrit [apsoly] en conformité avec cette règle de dévoisement, fournissant ainsi une représentation de la forme phonétique de surface. Cette position intermédiaire, de même qu'une certaine incohérence dans le choix de la représentation adoptée entre transcription phonologique et phonétique peut poser des problèmes, notamment pour une analyse des groupes de consonnes dans la langue française. En effet, il est clair que l'analyse du champ correspondant au mot 'médecin' fournira une information tronquée sur les groupes de consonnes car la représentation sous-jacente de ce mot ne contient pas de groupe consonantique. Du fait de la prise en considération de la chute de schwa, on est en mesure d'enregistrer la présence d'un groupe consonantique. Mais, si la représentation phonémique qui en est fournie dans la base de données en contient un, il n'est pas celui qui est effectivement réalisé par un locuteur natif. La transcription proposée dans cette base de données pour le mot 'médecin' ne correspond donc ni à la représentation sous-jacente, ni à la forme de surface.

Malgré ce problème, nous avons choisi d'utiliser BRULEX (Content et al., 1990) car c'était la seule base de données lexicale informatisée de la langue française qui nous permettait, du fait de la présence d'une transcription 'phonético-phonologique' des mots, de conduire l'analyse que nous souhaitions réaliser avec un outil informatique. Nous avons choisi d'intégrer dans notre analyse les groupes qui transgressent cette règle d'assimilation du trait de voisement en estimant que, même si leur occurrence peut refléter une observation incorrecte pour les groupes de consonnes pris dans leur individualité, ils fourniraient une information essentielle sur la fréquence d'occurrence des *catégories* de groupes de consonnes (occlusive-fricative ou occlusive-occlusive par exemple). Nous cherchions en effet à mettre en évidence une différence de fréquence entre des groupes tautosyllabiques (comme les occlusive-liquide) et hétérosyllabiques (comme les occlusive-fricative). Or ce problème de dévoisement de la consonne initiale concerne essentiellement les occlusive-fricative et les occlusive-occlusive, catégories dont nous souhaitions montrer la plus faible fréquence d'occurrence dans la langue

d'exploitation (une version libre de l'interpréteur AWK, gawk, peut être récupérée sur le site de la *Free Software Foundation* (http://www.gnu.org).

par rapport aux occlusives-liquides. Par conséquent, il nous a semblé important de tenir compte de la présence de ces groupes dans l'analyse puisque, transcription erronée ou pas du trait de voisement, leur appartenance à une classe phonétique n'est pas modifiée par la transcription de ce trait. Nous supposons par ailleurs que la quantité relativement importante de mots présents dans ce lexique devrait permettre d'obtenir des données statistiquement fiables malgré la présence de certains défauts dans la constitution de la base de données ; en effet, un comptage rapide des groupes de consonnes transcrits en [+ voisé][- voisé] aboutit à une quantité relativement faible. Dans BRULEX (Content et al., 1990), on recense ainsi 64 mots contenant une séquence de constrictives occlusives ou fricatives dont la forme est [+ voisé][- voisé] alors que la base de données contient au total 18608 mots intégrant un groupe de consonnes quel qu'il soit. Par ailleurs, il apparaît qu'une majorité de ces 64 mots sont des mots composés (par exemple 'protège-cahier'), ce qui justifie la transcription adoptée puisque la règle de dévoisement peut ne pas s'appliquer dans cette situation.

1.2.2. Base de données de mots isolés

Il est par ailleurs essentiel de garder à l'esprit que l'analyse des groupes de consonnes que nous allons proposer ici est en partie biaisée par le choix du corpus de langue que nous avons choisi. Cette base de données est constituée de mots isolés pour lesquels on dispose d'une information concernant la fréquence d'usage dans un corpus de langue. Il nous est cependant impossible d'identifier dans cette base de données des groupes de consonnes qui seraient générés par la juxtaposition de mots de la langue. Ainsi, dans la séquence 'une petite voiture' (/ynpətitvwatyr/), l'association de 'petite' et 'voiture' donne lieu à la prononciation du groupe de consonnes /tv/. Ce groupe apparaît dans un seul mot de la base de données sélectionnée ('tâte-vin'). Il est probable qu'une base de données de phrases aurait permis d'observer certains groupes de consonne en quantité plus importante, voire de faire émerger des groupes qui sont absents du corpus choisi. Un travail précédent nous conduit cependant à affirmer que ceci ne changerait probablement pas beaucoup le rapport des fréquences entre les divers groupes de consonnes étudiés. En effet, Malécot (1974) a conduit une analyse assez similaire à la nôtre sur un corpus de français parlé obtenu à partir d'enregistrements de conversations. Les données présentées ne sont pas assez détaillées pour nous permettre de les utiliser dans notre travail mais elles ont l'avantage de comparer la fréquence d'utilisation de chaque groupe de consonnes d'une part à l'intérieur des mots utilisés par les locuteurs et d'autre part aux frontières entre les mots. Cette comparaison permet à Malécot (1974) de mettre en évidence des fréquences moyennes très similaires pour les groupes recensés dans les mots utilisés et pour les groupes qui sont générés par la juxtaposition de ces mots dans les énoncés produits. Ainsi, la fréquence des groupes de consonnes recensés par Malécot (1974) reflète selon lui une tendance à utiliser prioritairement *à la frontière entre les mots* de la langue des groupes de consonnes qui sont plus fréquents *à l'intérieur des mots* de la langue. On peut donc s'appuyer sur ces données pour affirmer que les données présentées ici sont assez proches de celles qui auraient été obtenues avec un corpus de phrases.

2. Probabilité d'occurrence indépendante de la position

La première partie de l'analyse distributionnelle a consisté à estimer la fréquence des divers groupes de consonnes dans un lexique français sans prendre en compte leur position d'apparition dans les mots. Trois types de données ont été dérivées de cette analyse. La première catégorie de résultats fournit des informations sur le nombre de mots contenant un groupe de consonnes donné. Les deux autres ensembles de données pondèrent la fréquence de ces séquences de phonèmes par la fréquence des mots dans lesquels ils apparaissent (probabilité pondérée) ou par la fréquence de la première consonne (probabilité transitionnelle). Notre objectif est de montrer que des effets interprétables en termes de segmentation du signal de parole fondée sur les connaissances phonologiques des auditeurs mais reposant sur des processus cognitifs tout à fait différents pourraient expliquer les données obtenues par McQueen (1998) et Vroomen & De Gelder (1999). Si les données statistiques mettent en évidence un lien entre légalité / tautosyllabicité et fréquence, cette analyse nous mettra en position de réinterpréter les données présentées. En effet, si la syllabation des séquences de deux phonèmes peut se refléter dans leur fréquence d'apparition, il est alors possible de donner une interprétation alternative des données obtenues (en termes de segmentation probabiliste notamment, cf. section 4 de ce chapitre pour une discussion plus approfondie). C'est dans le but d'estimer l'impact de la fréquence d'occurrence des séquences de consonnes dans l'environnement linguistique sur les processus perceptifs que nous avons décidé de ne pas nous limiter à une information concernant le nombre de mots dans lesquels chaque séquence phonémique est attestée. En effet, si des processus purement probabilistes peuvent expliquer les effets obtenus, il est essentiel d'obtenir une approximation correcte de la fréquence avec laquelle une séquence de phonèmes est entendue dans les situations de communication naturelles. Or certains mots sont utilisés beaucoup plus fréquemment que d'autres. Il est donc essentiel, pour estimer la fréquence d'apparition d'une forme dans la langue, de pondérer le nombre de mots dans lesquels elle apparaît par leur fréquence d'usage. Le choix de réaliser une analyse des probabilités transitionnelles d'apparition est quant à lui dicté par des propositions récentes concernant les processus d'acquisition du langage ; propositions consistant à affirmer que le système cognitif de l'enfant mettrait en place une analyse des probabilités transitionnelles des séquences pour découvrir les frontières entre les mots avant d'avoir acquis un lexique (Saffran, Aslin, & Newport, 1996; Brent, 1996) et que l'adulte pourrait continuer d'avoir recours à ce type de procédures (Saffran, Newport et al., 1996; Brent & Cartwright, 1996; Brent, 1997) pour prédire la localisation des frontières de mots une fois ce lexique constitué. La notion de probabilités transitionnelles de même que sa distinction d'avec un simple calcul de fréquence est présentée dans la section 2.1.3. Elle consiste à tenir compte de la fréquence de la première consonne pour estimer la probabilité d'apparition de la seconde consonne.

2.1. Méthode d'analyse

Les analyses statistiques ont été effectuées sur ordinateur à l'aide du langage de programmation AWK (Aho et al., 1988). Pour extraire les ensembles de données présentés dans cette section, le script 'lisait' un fichier contenant la liste des séquences de phonèmes à rechercher dans le lexique (la base de données). Une fois cette liste recensée, il parcourait les champs de la base de données correspondant à la transcription phonémique et recherchait les occurrences de chacune des séquences qui lui étaient fournies en entrée sans se préoccuper de leur position dans les mots. Pour chaque séquence rencontrée dans la base, il incrémentait deux variables : l'une correspondait au nombre de mots contenant cette séquence, l'autre correspondait à la fréquence d'usage des mots qui est fournie dans BRULEX (Content et al., 1990). A la fin de l'analyse, le script sauvegardait un fichier texte qui contenait la liste des séquences de phonèmes recherchées et pour chacune d'entre elles le nombre de mots contenant cette séquence assorti de leur fréquence cumulée. Parallèlement à chaque analyse, nous avons également recensé les occurrences de consonnes prévocaliques afin d'estimer la fréquence des suites CV et de la comparer à celle des suites CC (le script adéquat pour le recensement du nombre de mots constitués de ces séquences est reproduit en Annexe 4, p.IV).

2.1.1. Fréquence d'occurrence

A partir des informations fournies par le script d'analyse (celui-ci est reproduit en Annexe 2, p.III), nous disposons directement d'une première série de données. Cette information est appelée ici *fréquence d'occurrence* des groupes de consonnes. Elle correspond au nombre de mots dans lesquels ces groupes apparaissent.

2.1.2. Probabilités pondérées

En second lieu, nous présentons les résultats d'une analyse de la fréquence de ces groupes en la pondérant par la fréquence des mots dans lesquels ils apparaissent. Nous appelons cet indice une *probabilité d'occurrence* dans la langue. Ces données fournissent une estimation plus fiable de la fréquence d'apparition des séquences dans les situations de communication linguistique puisqu'elles prennent en considération la fréquence d'usage des mots dans lesquels les groupes de consonnes sont prononcés. On peut en effet envisager qu'un groupe de consonnes soit relativement rare mais qu'il soit utilisé dans des mots qui, quant à eux, sont très souvent utilisés dans la langue. Ce groupe rare aurait alors, en réalité, une fréquence d'apparition assez élevée dans la langue malgré le nombre limité de mots qui le contiennent. Cette analyse a été limitée aux mots pour lesquels il existe, dans BRULEX (Content et al., 1990), une information sur la fréquence d'usage. Un second script a donc été utilisé ici qui se restreignait à analyser les entrées de la base de données pour lesquelles une information sur la fréquence du mot était fournie (le script correspondant est reproduit en Annexe 3, p.IV). La pondération a été réalisée en multipliant le nombre de mots dans lesquels le groupe apparaît par la fréquence cumulée de l'ensemble de ces mots. Du fait des différences de valeur importantes entre les résultats obtenus pour les groupes rares et ceux obtenus pour les groupes très fréquents, le résultat de cette multiplication à été transformé en valeurs logarithmiques décimales selon la formule suivante :

$$10 * \log_{10}(\text{Nombre\_de\_mots} * \text{Fréquence\_cumulée})$$

La multiplication par 10 a simplement pour objet de générer des valeurs approximées entières. Cette transformation logarithmique nous semble faciliter la visualisation de la distribution des probabilités moyennes d'occurrence des divers groupes de consonnes étudiés car elle permet de présenter sur une échelle de dimension restreinte les données correspondant aux séquences rares aussi bien que celles liées aux séquences fréquentes.

2.1.3. Probabilités transitionnelles

Finalement, le troisième type de données qui sont présentées ici prend en compte la fréquence de la première consonne du groupe pour estimer les probabilités transitionnelles des séquences qui nous intéressent. Dans l'estimation de la fréquence d'une séquence XY, une valeur élevée peut parfois s'expliquer essentiellement par une haute fréquence d'occurrence de la forme X. Si X est très fréquent, il apparaît souvent dans le corpus ; la forme Y a alors beaucoup de chances d'apparaître en conjonction avec la forme X. Cependant, cette valeur élevée de la fréquence d'apparition de XY est en partie déterminée par la fréquence de X. Cette fréquence de la suite XY peut donc, paradoxalement, n'être pas très informative.

Dans le cadre de la théorie de l'information (Shannon, 1948), l'informativité est en effet une notion très spécifique. Une forme est informative si sa présence est en mesure d'apporter une quantité d'information plus importante que celle dont elle est intrinsèquement constituée. Notamment, une forme qui permet de prédire la présence d'autres formes est très informative car elle est en mesure de fournir des informations permettant par exemple d'identifier un signal malgré une dégradation de celui-ci liée au médium de transmission. Une séquence XY fréquente n'est pas nécessairement informative. Si X est fréquent, il peut être suivi de n'importe quelle forme (aussi bien Y que Z). Par contre, si la présence de X permet de prédire avec un niveau de certitude élevé la présence de Y mais pas celle de Z, alors on peut dire que X est informatif. C'est dans ce cas précis que la suite XY aura un taux de probabilité transitionnelle élevé. Supposons deux séquences AB et CD. AB est fréquente alors que CD est rare. Il est possible que ces deux séquences ne soient pas plus informatives l'une que l'autre. Si A est très fréquent et que C est très rare, la présence de C peut éventuellement permettre, autant que celle de A, de prédire la forme suivante. C'est par exemple le cas si, malgré une fréquence d'occurrence très restreinte de C, la suite CD est relativement fréquente par rapport aux autres suites contenant C, c'est à dire si la présence de C détermine avec une certitude élevée celle de D. Si l'on se réfère à la théorie de l'information introduite par Shannon (1948), ce n'est donc pas la fréquence en soi qui est informative mais le rapport entre fréquence de la forme globale et celle de l'une des deux formes.

Nous avons insisté, dans le Chapitre 2, sur le caractère sériel (au moins en partie) du traitement appliqué par le système cognitif sur ce signal. Il nous semble donc logique de conduire cette analyse en termes de probabilités transitionnelles en prenant comme référent du calcul des probabilités transitionnelles la première consonne du groupe. Dans le cadre d'une comparaison des groupes de consonnes entre eux, cette distinction entre fréquence et probabilité transitionnelle est importante si l'on compare des groupes de consonnes qui commencent par une consonne différente. En effet, tant que l'on compare des groupes de consonnes qui partagent un élément commun, une estimation de la probabilité transitionnelle des séquences n'apporte rien de plus qu'une analyse de la fréquence de ces séquences dans le corpus puisque la fréquence de l'élément initial est la même. Par contre, lorsque l'on comparera des groupes de consonnes qui ne partagent pas la même consonne initiale, il sera important de recourir à cet indice de probabilité transitionnelle pour estimer leur différence en termes probabilistes.

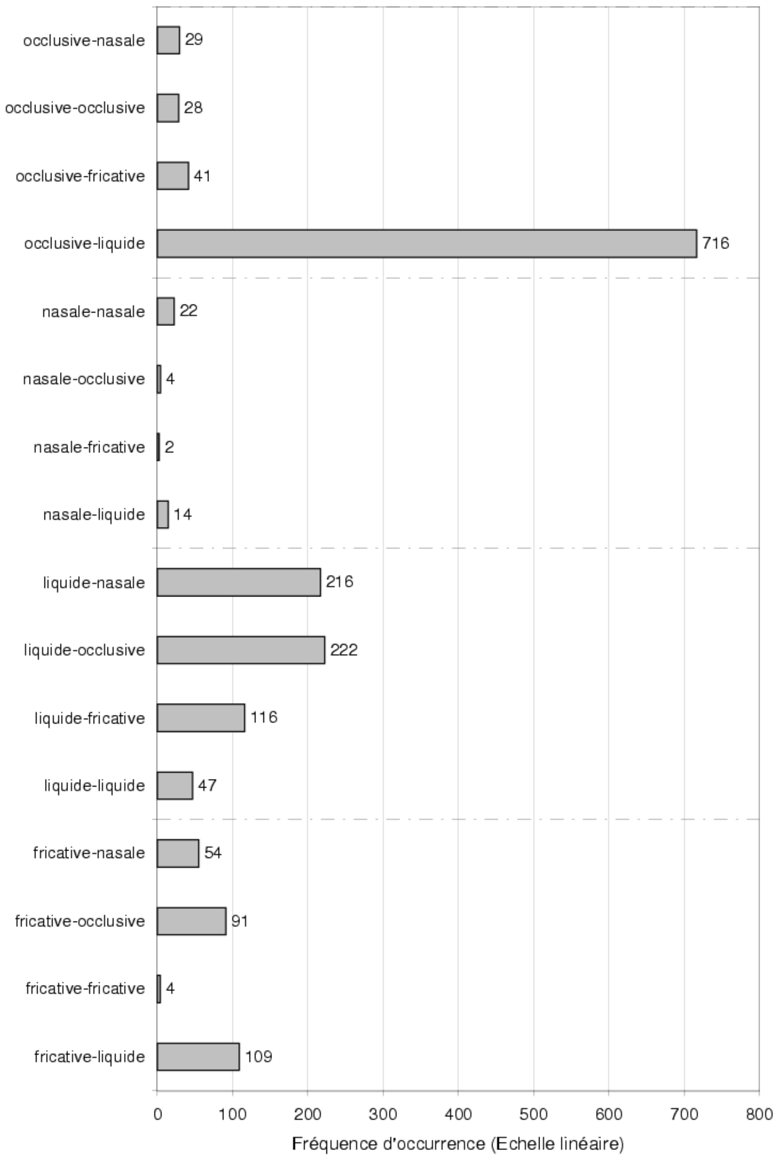

Figure 18 : Nombre moyen d'occurrences des groupes de consonnes en fonction de leur structure phonémique (mode d'articulation). Données calculées sur le corpus BRULEX (Content et al., 1990).

2.2. Résultats

Nous présentons en premier lieu les moyennes obtenues par catégorie de groupe en fonction du mode d'articulation de chaque consonne (occlusive, fricative, nasale, liquide). Nous restreignons notre description à un certain nombre de groupes qui apportent des informations particulièrement intéressantes en ce qui concerne le lien entre syllabation et fréquence. Les données intégrales pour chaque groupe sont présentées en Annexe 5 (p.V). Dans une seconde étape, nous étudions la distribution des fréquences individuelles de chaque groupe en fonction de sa catégorie phonétique.

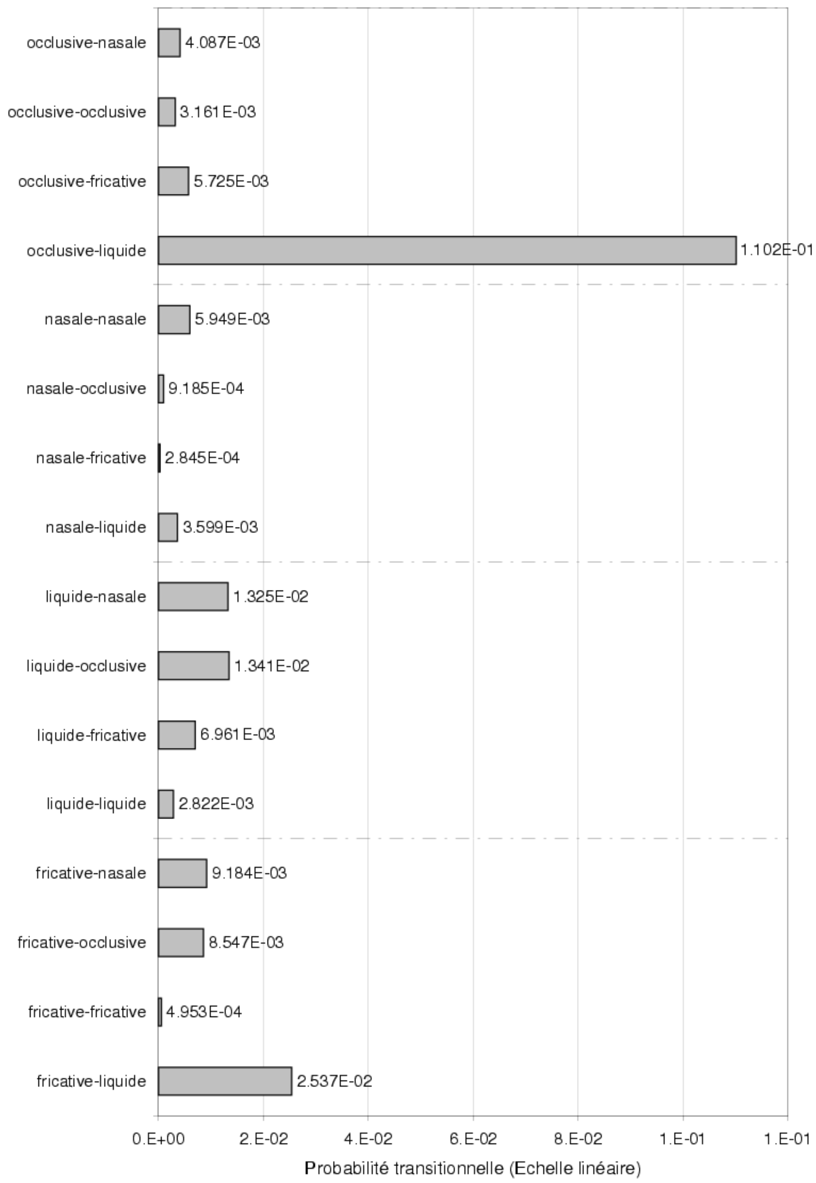

Figure 19 : Probabilité transitionnelle moyenne des groupes de consonnes en fonction de leur structure phonémique (mode d'articulation). Les fréquences du groupe de la consonne initiale sont calculées à partir de la base de données BRULEX (Content et al., 1990).

Une première observation importante est que les trois catégories de calculs que nous avons effectuées sur les groupes de consonnes présentent une forte corrélation positive entre elles. Ainsi, la fréquence des groupes de consonnes (leur nombre d'occurrences) et la fréquence d'usage des mots qui les contiennent sont corrélées positivement (r = .97). Cette corrélation met en évidence une tendance, pour les groupes de consonnes fréquents, à être utilisés dans des mots également fréquents. On observe le même phénomène pour le calcul de la corrélation entre fréquence et probabilité transitionnelle (r = .83) ; ce qui implique une homogénéité de l'utilisation des consonnes dans les divers groupes recensés. On peut ainsi raisonnablement

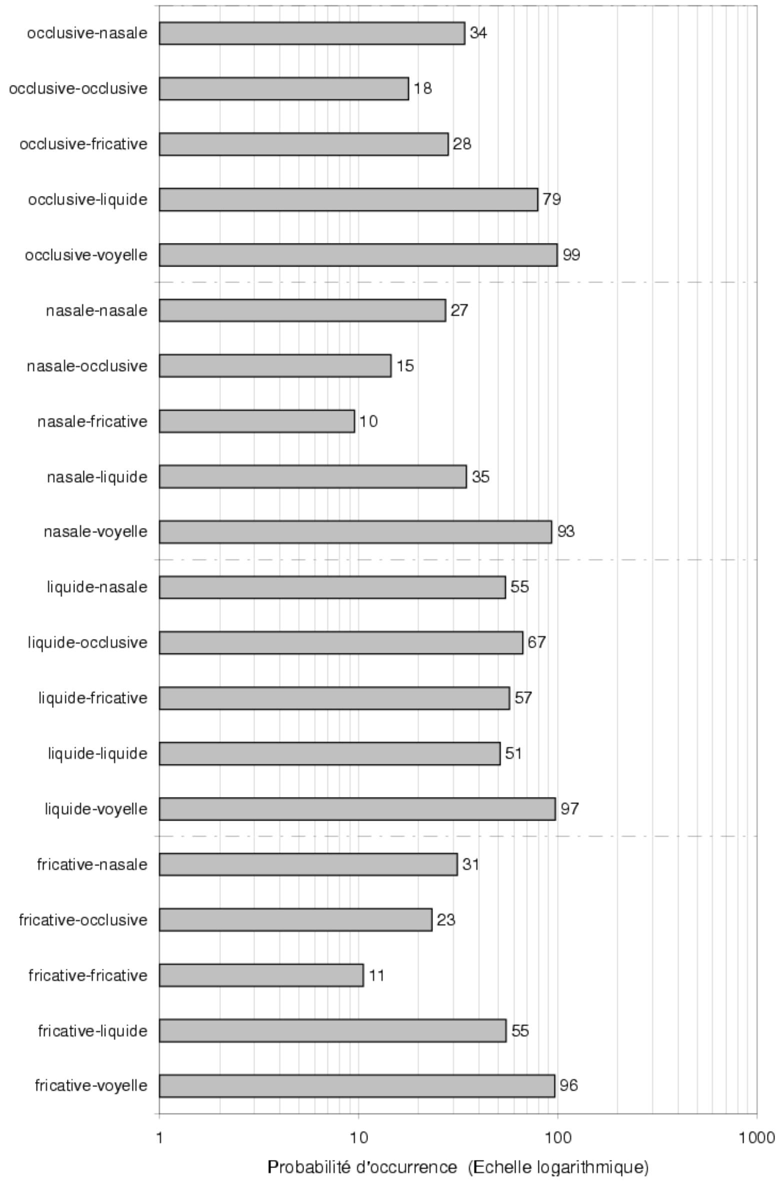

Figure 20 : Probabilité d'occurrence moyenne des groupes de consonnes en fonction de leur structure phonémique (mode d'articulation). Données calculées sur le corpus BRULEX (Content et al., 1990).

affirmer que l'indice de probabilité transitionnelle n'apporte pas beaucoup plus d'information que celui de la fréquence d'occurrence dans le lexique, du moins pour l'étude des fréquences moyennes par catégorie phonétique. La valeur du coefficient de corrélation est cependant moins élevée qu'entre fréquence d'occurrence et fréquence d'usage des mots, ce qui indique qu'une quantité plus importante de groupes de consonnes donne lieu à une non-correspondance entre fréquence et probabilité transitionnelle qu'entre fréquence et probabilité d'occurrence.

2.2.1. Moyennes

La première étape de notre analyse consiste à étudier la fréquence moyenne (selon les 3 indices décrits précédemment) des groupes de consonnes en fonction de leurs caractéristiques phonétiques de mode d'articulation. Les graphiques suivants illustrent les différences de fréquence moyenne entre catégories. Dans le premier (Figure 18) sont présentées les données moyennes obtenues pour le comptage du nombre de mots dans lesquels chaque groupe consonantique est recensé. Nous ne présentons pas dans ce graphique les données obtenues pour les séquences CV (Consonne-Voyelle). Cette omission nous permet de mieux visualiser les différences de fréquence entre groupes de consonnes, ceux-ci étant toujours nettement plus rares que les séquences CV.

Les deux autres graphiques reproduisent respectivement les probabilités transitionnelles (Figure 19) et les probabilités d'occurrence (fréquence du groupe pondérée par la fréquence d'usage des mots, Figure 20). Ce dernier permet de présenter, outre les résultats observés pour les groupes de consonnes, les données obtenues pour les séquences Consonne-Voyelle. Le choix d'une transformation logarithmique nous permet en effet de représenter simultanément les données correspondant aux séquences très rares et très fréquentes. On remarquera ici le cas particulier des groupes fricative-nasale et fricative-occlusive dont la position sur l'échelle du nombre d'occurrences (Figure 18) est l'inverse de celle observée sur l'échelle des probabilités pondérées (Figure 20) ; ceci met en évidence, malgré la forte corrélation observée entre nombre d'occurrences et fréquence d'usage des mots, la possibilité que cette fréquence d'usage puisse contribuer à la probabilité d'occurrence des suites de phonèmes dans la langue. Les données de probabilité transitionnelle ne diffèrent par contre pas des données de fréquence d'occurrence. Cette équivalence des résultats observés est probablement liée à l'équiprobabilité des phonèmes individuels dans la langue. Les résultats présentés seront donc désormais fondés sur les données de *probabilité d'occurrence* (fréquence d'occurrence pondérée par la fréquence d'usage des mots) car ils permettent à la fois de prendre en compte la productivité des groupes de consonnes dans le lexique et de comparer les groupes de consonnes avec les suites Consonne-Voyelle.

Du fait du nombre considérable de catégories à prendre en compte dans une comparaison statistique des moyennes observées en fonction du mode d'articulation des phonèmes constituant ces groupes de consonnes, nous avons conduit cette analyse à l'aide de tests statistiques *post-hoc*. La valeur élevée du nombre de degrés de liberté de la comparaison globale, ce nombre de degrés de liberté étant lié à la quantité d'observations, nous a conduit à sélectionner un test statistique relativement conservateur afin de limiter le nombre de comparaisons qui feraient émerger une différence significative (Winer, 1971). Nous avons donc utilisé le test de Scheffé pour comparer les moyennes entre elles. Nous focalisons notre analyse sur les groupes constitués d'une fricative ou d'une occlusive à l'initiale car ils nous semblent constituer deux catégories particulièrement propices à l'étude du lien entre syllabation et fréquence et qu'ils fournissent par ailleurs une quantité de données plus substantielle que les autres types de groupes consonantiques. Certains des membres de ces catégories peuvent en outre être raisonnablement considérés comme des attaques de syllabe bien formées. C'est en particulier le cas de la plupart des occlusive-liquide et fricative-liquide. Les groupes à initiale liquide ou nasale nous semblent plutôt correspondre à des coda syllabiques (/rk/ comme dans /bark/) ou même à des séquences comportant une frontière syllabique médiane (/nk/ comme dans /mankɛ̃/, la seule exception nous semblant être l'emprunt 'round'). Or la description de codas complexes (contenant plusieurs consonnes) n'est peut-être pas justifiée (Dell, 1995). Ce dernier propose au contraire que les groupes de consonnes décrits comme des codas complexes constituent plutôt une juxtaposition de deux positions syllabiques différentes, la consonne initiale du groupe ayant le statut effectif de coda alors que la seconde correspondrait à une attaque de syllabe (laquelle comporterait un noyau vide). La problématique que nous avons choisi d'aborder dans ce travail se fonde principalement sur des groupes de consonnes qui peuvent ou non être regroupés à l'attaque syllabique. Nous cherchons donc à déterminer ce qu'est une attaque de syllabe bien formée (ou légale). En restreignant notre analyse à deux catégories de groupes parmi lesquelles une part non négligeable (mais pas l'ensemble) peut être caractérisée comme une attaque de syllabe bien formée, nous serons en mesure d'utiliser ces données fréquentielles pour réanalyser les données comportementales auxquelles nous nous intéressons.

Les séquences C[32]-liquide présentent une probabilité moyenne d'occurrence plus élevée (79 pour les groupes à initiale occlusive ; 55 pour ceux à initiale fricative) que les autres catégories (respectivement 18, 28 et 34 pour les groupes à initiale occlusive ; 11, 23, 31 pour les groupes à initiale fricative). Les séquences C-voyelle présentent quant à elles des probabilités d'occurrence plus importantes que l'ensemble des autres catégories (respectivement 99 et 96 pour les occlusive-voyelle et les fricative-voyelle). L'analyse statistique que nous avons conduite met en évidence, comme il était possible de s'y attendre au vu des graphiques ci-dessus, des différences significatives de probabilité d'occurrence entre les différents groupes de consonnes. Le Tableau 3 présente les seuils de probabilité obtenus avec le test de Scheffé pour la comparaison des séquences commençant par une occlusive (Tableau 3a) et par une fricative (Tableau 3b). Nous avons tracé dans chacun de ces tableaux un rectangle indiquant la frontière

[32] C étant ici mis pour fricative ou occlusive.

entre les séquences qui sont essentiellement tautosyllabiques (C-voyelle, C-liquide) et celles qui seraient plutôt hétérosyllabiques (C-fricative, C-occlusive, C-nasale). A l'intérieur de ce rectangle, on trouve les seuils de probabilité obtenus en comparant des ensembles de séquences en général tautosyllabiques d'une part et hétérosyllabiques de l'autre. Les seuils de probabilité situés à l'extérieur de ce rectangle correspondent quant à eux à des comparaisons de séquences à l'intérieur de l'ensemble considéré comme tautosyllabique (à gauche) ou hétérosyllabique (en dessous du rectangle).

Tableau 3 : Seuils de probabilité des tests de Scheffé appliqués à la comparaison de probabilités d'occurrence des groupes de consonnes à initiale occlusive (a) ou fricative (b). Les comparaisons sont effectuées par catégorie de groupe en fonction du mode d'articulation des phonèmes. Les seuils de probabilité statistiquement significatifs sont retranscrits en caractères gras et italique. Les intitulés des lignes et des colonnes correspondent au second phonème de la séquence.

a/ Occlusive initiale	liquide	fricative	occlusive	nasale
voyelle	0.516	***0.000***	***0.000***	***0.000***
liquide		***0.000***	***0.000***	***0.000***
fricative			0.416	0.929
occlusive				0.171

b/ Fricative initiale	liquide	fricative	occlusive	nasale
voyelle	***0.010***	***0.000***	***0.000***	***0.000***
liquide		***0.000***	***0.002***	0.090
fricative			0.204	***0.037***
occlusive				0.814

Conformément à nos prédictions, il semble exister un lien entre la structure phonologique des groupes de consonnes et leur fréquence d'occurrence dans la langue. Ce lien est particulièrement prégnant si l'on observe les différences de fréquence dans la catégorie des groupes commençant par une occlusive. Il est clair que la plupart des groupes occlusive-liquide, de même que toutes les séquences occlusive-voyelle, sont produits en attaque syllabique et que, pour la plupart, aucune frontière syllabique / phonotactique n'est insérée entre les deux consonnes de ces groupes. Or ceux-ci sont en moyenne beaucoup plus fréquents que toutes les autres catégories de groupes consonantiques commençant par une occlusive. On observe par ailleurs une nette dissociation entre les deux catégories que nous considérons comme essentiellement tautosyllabiques (occlusive-liquide et occlusive-voyelle) et celles dont nous jugeons qu'elles correspondent plutôt à des groupes hétérosyllabiques (occlusive-fricative, occlusive-occlusive, occlusive-nasale). En effet, tous les seuils de probabilité du test de Scheffé sont inférieurs à 0.01

à l'intérieur du rectangle délimitant la frontière entre ces deux ensembles alors qu'aucune des comparaisons intra-catégorie n'est significative. Le test de Scheffé a également été conduit sur les données de fréquence d'occurrence et de probabilité transitionnelle. A l'exception des données de fréquence obtenues pour les groupes à initiale occlusive, il n'émerge que peu d'effets significatifs de ces comparaisons (cf. Annexe 6 ; p.XII). Pour la plupart des comparaisons, seule émerge une distinction entre les suites CV et CC[33].

Les données statistiques obtenues pour les groupes à initiale fricative (Tableau 3b) fournissent des résultats moins clairs concernant le lien entre tautosyllabicité et fréquence dans la langue. Il n'apparaît pas de dissociation nette entre les séquences fricative-voyelle et fricative-liquide d'une part et les 3 autres catégories d'autre part. Les suites fricative-liquide sont significativement moins fréquentes que les suites fricative-voyelle. Ceci ne contredit pas notre hypothèse puisque tautosyllabicité et fréquence pourraient tout à fait être liées sans induire une absence de différence fréquentielle intra-catégorie. On observe cependant, à l'intérieur du rectangle représentant la distinction entre groupes supposés tautosyllabiques et groupes supposés hétérosyllabiques, des comparaisons dont le test de Scheffé ne permet pas d'affirmer qu'elles correspondent à des séquences différant significativement en termes de fréquence d'occurrence. Il est probable que cette distinction moins nette entre les deux ensembles définis *a priori* soit liée à une plus grande variabilité dans la syllabation des groupes de consonnes commençant par une fricative que dans celle des groupes à initiale occlusive. Nous avons mentionné le caractère illégal des deux séquences /tl/ et /dl/ qui, malgré leur prise en compte dans les données des occlusive-liquide ne peuvent pas être tautosyllabiques. L'ensemble des groupes à initiale fricative contient certainement une quantité plus importante de groupes ne pouvant pas être classés simplement en fonction de leur mode d'articulation. Les groupes fricative-liquide comme /vr/ ou /fl/ constituent des attaques syllabiques bien formées. On trouve par exemple des suites fricative-liquide dans 'vrai', 'frelon', 'flibustier'… On trouve par contre dans cette catégorie des séquences qui pourraient être décrites comme hétérosyllabiques. C'est par exemple le cas de /sl/ qui, bien qu'il puisse constituer une attaque syllabique bien formée (comme dans 'slalom')

[33] Il est probable que ce soit lié à des différences de fréquence trop importantes entre les séquences CV et CC lorsque l'on utilise une échelle linéaire. En effet, la plupart des tests statistiques paramétriques a pour principe de comparer la taille de chaque effet en utilisant comme mesure de l'*erreur* la variance globale des mesures effectuées. Dans le cas d'une échelle linéaire des fréquences (ou des probabilités transitionnelles), Les suites CV ont une valeur considérablement plus importante que les suites CC sur cette échelle. Elles accroissent donc la variance à un tel point que le test n'est plus en mesure de détecter que des différences moyennes très importantes. Le choix d'une échelle logarithmique pour la présentation des probabilités d'occurrence a certainement permis de réduire la variance introduite par les séquences CV et de faire émerger, de fait, ces différences entre catégories de groupes consonantiques.

pourrait être considéré comme hétérosyllabique en position intervocalique (dans ‘islam’ par exemple). On peut certainement observer ce type de variabilité parmi les autres catégories de groupes. Le cas des fricative-occlusive est particulièrement frappant. Les groupes /s/ + occlusive apparaissent régulièrement en début de mot ; ce qui n’est pas le cas des groupes /f/ + occlusive. Les premiers pourraient constituer des attaques de syllabe bien formées. Il est par conséquent probable que les catégories comparées pour les groupes à initiale occlusive présentent une variabilité plus limitée que celles correspondant aux groupes à initiale fricative. Cette variabilité plus restreinte pourrait stabiliser les résultats statistiques obtenus. Malgré cette relative incohérence des données statistiques obtenues pour les groupes à initiale fricative, on retrouve dans le tableau des seuils de probabilité du test de Scheffé une certaine tendance à grouper les séquences tautosyllabiques et hétérosyllabiques ensemble : parmi les 6 comparaisons effectuées entre les catégories constituant ces deux ensembles, cinq sont statistiquement significatives. Cette proportion nous semble fournir un assez bon indicateur du lien que nous cherchons à mettre en évidence. L’analyse des groupes à initiale fricative contribue donc également, bien que dans une moindre mesure, à l’affirmation d’un lien entre syllabation et fréquence.

Ces données moyennes ne fournissent cependant pas une explication satisfaisante du lien qu’il est possible de décrire entre fréquence et syllabation. En effet, la question qui nous semble essentielle ici est de déterminer si, alors que fréquence et tautosyllabicité sont effectivement corrélées, il est possible de dissocier ces deux variables ou si, au contraire, elles constituent deux expressions d’un seul et même phénomène : une régularité phonologique observable déterminée par les contraintes inhérentes à la langue et dont ces deux manifestations seraient indissociables. Il est déjà possible de prédire, à partir des données statistiques fournies par le test de Scheffé, que tel n’est pas le cas. En effet, on observe pour les séquences à initiale fricative des incohérences dans la distribution des seuils de significativité qui ne permettent pas entièrement de conclure à une correspondance terme à terme entre tautosyllabicité et fréquence. Cette incohérence pourrait en partie s’expliquer par l’intégration, dans chaque ensemble (tautosyllabique vs. hétérosyllabique), de groupes de consonnes qui appartiennent sans aucun doute à l’ensemble opposé. On peut cependant affirmer que, même si nous avions regroupé dans le cadre de cette analyse des groupes ‘purs’ dont il serait impossible de dire qu’ils devraient être classés dans l’ensemble opposé, nous aurions pu observer un certain flou dans les résultats statistiques. La première raison est liée au caractère conservateur du test de Scheffé qui présente plutôt une tendance à ne pas rejeter l'hypothèse nulle[34]. Il reste néanmoins que certaines comparaisons font émerger une différence significative à laquelle nous ne nous attendions pas. Ceci peut en partie être lié aux multiples comparaisons qui sont effectuées dans le cadre d'un test post-hoc, cette quantité importante aboutissant au risque d'observer par hasard des effets significatifs même si les tests *post-hoc* sont justement conçus pour limiter ces risques. Une seconde explication -laquelle pourrait rendre compte de l'émergence d'effets significatifs inattendus sans reposer sur des considérations purement statistiques- est que, peut-être, fréquence et tautosyllabicité constituent des phénomènes corrélés mais sans réelle source d'explication commune. Notamment, il est possible que la tendance à utiliser plus fréquemment des groupes tautosyllabiques dans les mots de la langue ne soit que partiellement déterminée par les contraintes phonologiques de la langue ; d'autres paramètres pouvant alors intervenir (comme le hasard). Ces deux observables pourraient alors constituer des phénomènes corrélés mais en partie dissociables. Si c'est le cas, il est alors certainement possible de mettre en évidence des recouvrements dans les distributions de probabilité d'occurrence catégorisée en fonction des types de groupes de consonnes ; recouvrements qui nous permettraient de séparer tautosyllabicité et fréquence et fourniraient par conséquent les moyens d'une étude plus précise de la contribution respective des régularités phonologiques et de la fréquence des groupes de phonèmes dans les résultats comportementaux obtenus McQueen (1998) et par Vroomen & De Gelder (1999).

2.2.2. Distributions

Si la différence de fréquence observée est directement liée à la tautosyllabicité des groupes de consonnes et que ces deux paramètres sont indissociables, alors il sera impossible de déterminer si des processus faisant intervenir des représentations phonologiques sont effectivement à l'œuvre dans l'émergence des effets observés par McQueen (1998) et Vroomen & De Gelder (1999) ou si ces effets pourraient au contraire être déterminés par des processus impliquant le recours à des calculs probabilistes tels que ceux proposés par Saffran, Newport, & Aslin (1996) ou par Brent & Cartwright (1996). Puisque la fréquence des groupes de consonnes dans la langue est fortement liée à leur syllabation, il est essentiel de déterminer si ces deux variables sont séparables ou si, au contraire, elles correspondent à deux facettes d'un phénomène unique. Si fréquence et tautosyllabicité sont liées mais pas confondues, leurs distributions devraient se chevaucher. La mise en évidence de cette orthogonalité des dimensions fournirait

[34] On notera que pour les groupes à initiale fricative, le seul seuil non-significatif dans le rectangle dissociant séquences tautosyllabiques et hétérosyllabiques pourrait en ce sens être considéré comme marginal (p=.090).

alors un outil permettant de tester indépendamment les effets respectifs de la fréquence et de la structuration syllabique dans les processus cognitifs de segmentation lexicale. Afin de répondre à cette question, nous présentons les graphiques correspondant aux distributions de probabilité d'occurrence des groupes de consonnes en les classant par catégorie phonétique de mode d'articulation. La Figure 21 illustre les distributions des divers groupes de consonnes commençant par une occlusive (/b/, /d/, /g/, /p/, /t/, /k/). Les données correspondant aux groupes à initiale fricative (/v/, /z/, /ʒ/, /f/, /s/, /ʃ/) sont présentées dans la Figure 22.

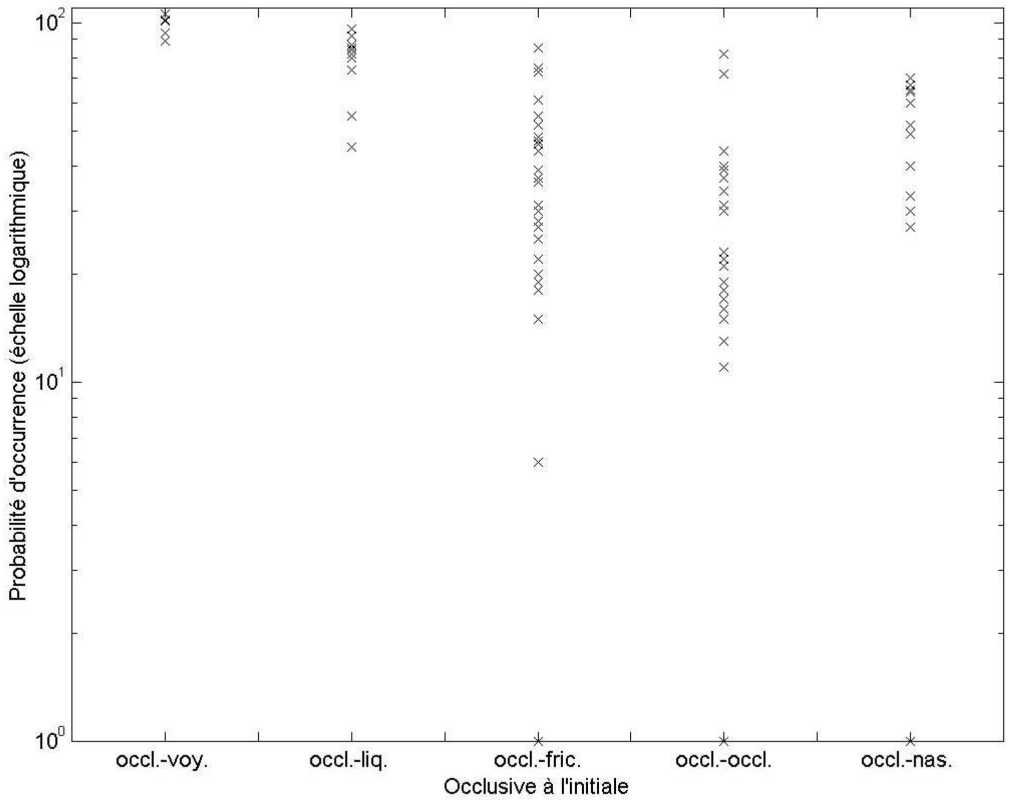

Figure 21 : Distribution des probabilités d'occurrence pour les groupes de consonnes à initiale occlusive.

On peut voir, malgré la tendance d'une fréquence plus élevée des groupes de consonnes tautosyllabiques mise en évidence dans la section précédente, que ces deux variables ne sont pas confondues. Que l'on observe aussi bien le graphique correspondant aux groupes à initiale occlusive que celui des distributions correspondant aux groupes à initiale fricative, les distributions de probabilité d'occurrence se recoupent ; ce qui signifie que deux groupes de consonnes ayant des caractéristiques phonétiques différentes peuvent présenter une fréquence similaire. Notamment, un groupe ayant tendance à être tautosyllabique peut être aussi fréquent qu'un groupe hétérosyllabique. De même, deux groupes de caractéristiques phonétiques similaires peuvent présenter des différences de fréquence de grande amplitude, qu'ils soient hétérosyllabiques ou pas. On remarque que cette observation vaut aussi bien pour les groupes à initiale occlusive que pour ceux à initiale fricative. Si l'on admet que ces groupes peuvent se scinder en deux catégories : les groupes fricative-liquide et occlusive-liquide appartenant pour la plupart à l'ensemble des groupes tautosyllabiques et les autres types de groupes (C-fricative, C-occlusive, et C-nasale) se prononçant essentiellement de manière hétérosyllabique, il existe -malgré des différences importantes de fréquence moyenne- des paires présentant une probabilité d'occurrence identique mais n'appartenant pas au même ensemble de groupes consonantiques.

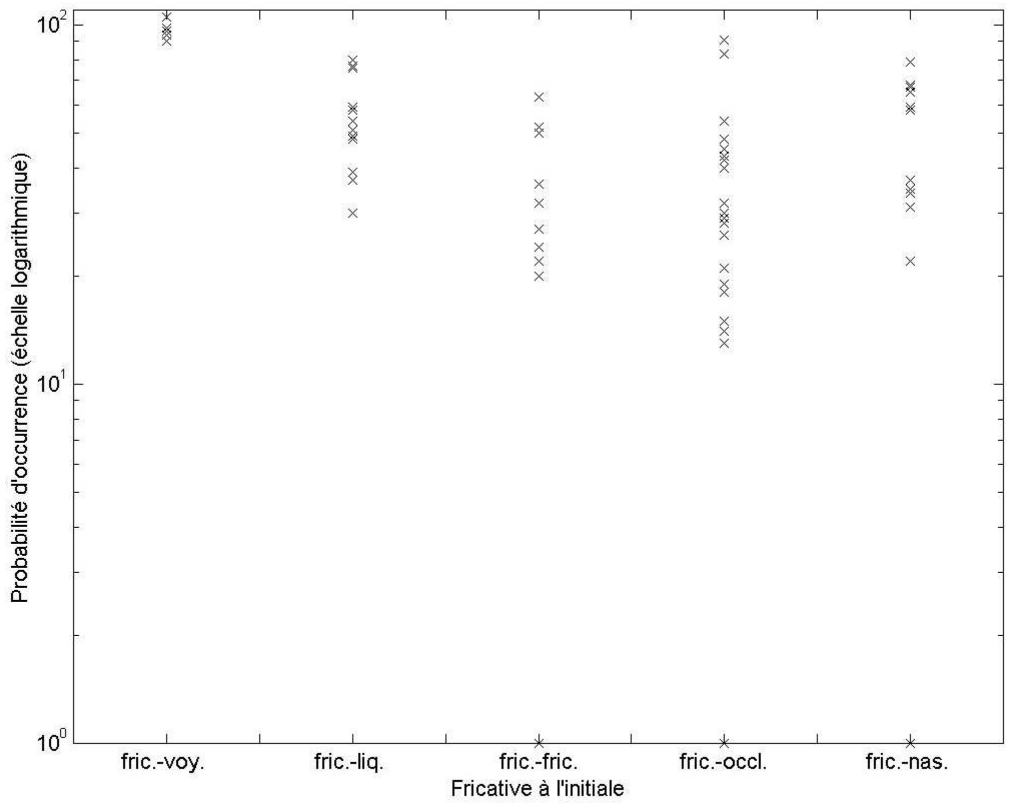

Figure 22 : Distribution des probabilités d'occurrence pour les groupes de consonnes à initiale fricative.

2.3. Discussion

Si fréquence et tautosyllabicité sont liées, elles ne sont cependant pas confondues. Il est par conséquent concevable de dissocier ces deux paramètres dans une étude du rôle de la structuration syllabique ou des contraintes phonotactiques dans les processus cognitifs de segmentation de la parole en mots. En effet, nous avons observé que les groupes typiquement prononcés à l'attaque de syllabe ont une tendance à apparaître plus fréquemment que les autres dans les mots de la langue. Ce résultat met en évidence une relation entre un observable (la tautosyllabicité) dont la cause supposée est le respect de contraintes phonologiques intégrées au système linguistique propre à la langue et un autre observable (la fréquence) qui pourrait se manifester indépendamment de contraintes linguistiques. Au contraire, la fréquence d'utilisation des séquences de consonnes quelle que soit leur position dans les mots n'est pas indépendante des principes de syllabation déterminant la structuration des consonnes dans la chaîne parlée. Nous verrons, dans la section 4 de ce chapitre que cette observation permet de réanalyser les données présentées par McQueen (1998) et Vroomen & De Gelder (1999) en proposant quelques interprétations alternatives. La section 2.2.2 nous a cependant permis de mettre en évidence une dissociation entre tautosyllabicité et fréquence. Cette dissociation apparaît avec l'observation des distributions de probabilité d'occurrence en fonction de la catégorie de mode d'articulation. Malgré la correspondance observée entre probabilité moyenne d'occurrence et caractère tauto- ou hétérosyllabique, l'analyse précédente permet d'observer un recouvrement des distributions de probabilité d'occurrence relevées pour les diverses catégories de groupes de consonnes. Cette observation vaut aussi bien pour les groupes à initiale fricative que pour ceux à initiale occlusive. Nous sommes donc en mesure d'affirmer que, si fréquence d'occurrence et tautosyllabicité des groupes de consonnes sont liées, elles ne sont pas confondues. De fait, il est légitime d'affirmer que la fréquence d'occurrence d'un groupe de consonnes est un observable qui n'est pas intégralement déterminé par les contraintes linguistiques de la langue. Il devrait donc être possible de dissocier les effets respectifs de la fréquence et de la structure phonologique (qu'elle soit liée à la syllabicité ou à la légalité phonotactique) dans des études portant sur les processus impliqués dans la segmentation de la parole en mots.

3. Probabilité d'occurrence en début de mot

Il est cependant nécessaire, du fait de cette absence de correspondance stricte entre tautosyllabicité et fréquence, de choisir un critère opérationnel qui nous permettra de catégoriser les groupes selon des critères purement phonologiques (phonotactiquement légal vs. illégal ou tautosyllabique vs. hétérosyllabique) et de distinguer celui-ci d'un critère probabiliste sur lequel serait fondée une catégorisation en classes fréquentielles. Pour cela, il nous a semblé intéressant de conduire une analyse similaire à la précédente mais qui serait restreinte à estimer la fréquence des groupes de consonnes en début de mot. Afin de décrire chaque groupe de consonne en termes de structuration syllabique, il nous faut adopter une définition opérationnelle de ce qu'est une attaque de syllabe bien formée. Or nous avons vu dans la description des principes de syllabation aussi bien que dans la section consacrée à la question de la légalité phonotactique que différentes conceptions de ces notions peuvent être rencontrées. Dans les travaux réalisés en phonologie, on considère en général que toute séquence attestée dans la langue en une position donnée (par exemple en début de mot) est phonotactiquement légale dans cette position (ou constitue une attaque syllabique bien formée). Si l'on se réfère au travail de Dell (1995), le fait qu'un seul exemplaire soit attesté dans la langue en début de mot conduit à le considérer comme une attaque de syllabe bien formée. Il nous semble cependant que ce critère n'est pas assez discriminant. En effet, si l'on recense l'ensemble des groupes attestés en position initiale de mot, on s'aperçoit qu'une quantité considérable de groupes de consonnes peut apparaître en position initiale dans au moins un mot de la langue.

Tableau 4 : Exemples de groupes de consonnes attestés en début de mot dans la base de données BRULEX (Content et al., 1990) mais qui seraient certainement hétérosyllabiques en position intervocalique.

Groupes consonantiques	Nombre d'occurrences	Notes	Exemples
/ps/	44	37 de la famille 'psy'	psychologie, pseudonyme
/dʒ/	9	emprunts récents	jazz
/ts/	7	emprunts récents	tsar, tsigane
/ft/	4	même famille	phtisie
/ʃn/	3	même famille	schnock
/mn/	2	même famille	mnémonique

Nous avons vu que la fréquence d'occurrence quelle que soit la position dans les mots n'est pas non plus un critère discriminant permettant de dissocier les séquences tautosyllabiques et hétérosyllabiques puisque les distributions de probabilité d'occurrence se recouvrent largement. La notion de déviance introduite par Dell (1995) pourrait être utile si l'on était en mesure de définir ce qu'est un groupe déviant. En effet, il suffirait de considérer uniquement les groupes attestés dans la langue mais non-déviants comme tautosyllabiques et les autres comme hétérosyllabiques pour constituer nos deux catégories. Cependant, l'absence de critère permettant d'affirmer qu'un groupe est 'légal mais déviant' rend difficile le choix de cette solution. En outre, si l'on admet comme critère de bonne forme en attaque syllabique le fait qu'une séquence soit attestée en position initiale de mot dans la langue, on est en mesure de recenser -même pour les groupes dits non-déviants (Dell, 1995)- un certain nombre de groupes qui se prononceront de manière hétérosyllabique en position intervocalique. Le Tableau 4 présente une sélection de groupes de consonnes qui sont attestés dans la langue mais doivent plutôt être considérés comme hétérosyllabiques. Bien que la plupart de ces mots constituent des emprunts récents ou soient utilisés assez rarement dans la langue, ils doivent être pris en compte afin de déterminer un critère de distinction entre groupes tautosyllabiques et hétérosyllabiques. Nous faisons l'hypothèse que la combinaison de critères probabilistes et positionnels (le fait qu'un groupe de phonèmes soit prononcé en début de mot) permettra de dissocier clairement ces deux types de groupes de consonnes.

Nous ne recherchons pas ici un critère qui nous permettrait de déterminer de manière binaire -pour autant que ce soit possible- quels groupes constituent des attaques syllabiques *bien formées* ou pas (c'est à dire quel groupe peut être prononcé en attaque syllabique et être considéré comme une *bonne forme* phonologique par un locuteur natif). Il est donc important de distinguer l'objectif que nous nous fixons de celui du phonologue qui se pose des problèmes tout à fait différents. Nous avons insisté, dans le cadre du Chapitre 3, sur les rapports et les distinctions qui peuvent exister entre légalité phonotactique et syllabation. Il est probable que la légalité phonotactique puisse se décrire en réalité comme un continuum allant du 'tout à fait légal' au 'totalement illégal' en attaque de syllabe, une quantité considérable de séquences phonémiques se situant entre ces deux extrêmes. Ce phénomène pourrait être lié à la variabilité qui est couramment observée concernant la syllabation des séquences de phonèmes en position intervocalique et aux difficultés rencontrées en phonologie pour décrire un modèle non-falsifié de ces procédures de syllabation. Notre objectif est d'aboutir à un critère opérationnel qui serait différent de la fréquence d'occurrence calculée indépendamment de la position dans les mots et qui permettrait de catégoriser les groupes de consonnes du français dans les classes tautosyllabique ou hétérosyllabique sur la base d'une mesure observable dans la langue. Cet indice pourrait s'apparenter à un *observable phonologique* et nous permettrait alors de distinguer indices probabilistes et phonologiques dans la constitution de nos expériences. Ce que nous souhaitons obtenir est donc une mesure *observable dans la langue* du taux de bonne formation des groupes de consonnes en attaque syllabique.

3.1. Méthodes d'analyse

Les mêmes catégories d'informations que celles de la section précédente peuvent être extraites pour une analyse de la fréquence des groupes de consonnes en début de mot. Nous nommons les indices dérivés ici des *mesures de fréquence positionnelles*. Par contraste, les données obtenues dans l'analyse précédente seront appelées des *mesures de fréquence brutes*. Seules les probabilités transitionnelles n'ont pas été estimées ici. En effet, il nous semble que si le système cognitif est en mesure d'utiliser des calculs probabilistes dans le cadre des processus de traitement de la parole, il serait peu efficace de faire reposer ces calculs sur des représentations élaborées (les mots, les syllabes, …). L'objectif que nous poursuivons ici est de proposer *un indice mesurable du taux de bonne formation en attaque syllabique*. Nous n'envisageons pas que ce type d'information (même s'il repose sur une mesure de fréquence) soit utilisé par le système cognitif dans le traitement qu'il effectue sur les stimuli qui lui parviennent. Il serait beaucoup plus économique d'utiliser une information fréquentielle pure. La méthode d'analyse est la même que dans la section précédente. Les scripts AWK ont été modifiés afin de restreindre la recherche des groupes de consonnes (et des séquences Consonne-Voyelle) aux débuts de mots (cf. Annexe 7, Annexe 8 et Annexe 9, pp.XIII-XIV).

3.1.1. Fréquence d'occurrence

A partir des informations fournies par le script d'analyse, nous disposons directement d'une première série de données : la *fréquence d'occurrence* des groupes de consonnes en début de mot, c'est à dire le nombre de mots dans lesquels chaque groupe apparaît en position initiale dans la base de données.

3.1.2. Probabilités pondérées

En second lieu, nous présentons les résultats d'une analyse de la fréquence de ces groupes en position initiale en la pondérant par la fréquence des mots dans lesquels ils apparaissent. De même que dans la section précédente, nous appelons cet indice une *probabilité d'occurrence* (en début de mot) dans la langue. Cette analyse a été limitée aux mots pour lesquels il existe, dans BRULEX (Content et al., 1990), une information sur la fréquence d'usage. La même formule de transformation des données en valeurs logarithmiques a été appliquée pour dériver les valeurs de probabilité d'occurrence en position initiale de mot :

$$10 * \log_{10}(\text{Nombre\_de\_mots} * \text{Fréquence\_cumulée})$$

Cette formule permet à nouveau de visualiser les fréquences des différentes séquences sur une échelle facilitant leur comparaison.

3.2. Résultats

Nous présentons en premier lieu les moyennes obtenues par catégorie de groupe en fonction du mode d'articulation de chaque consonne (occlusive, fricative, nasale, liquide). Nous restreignons notre description aux groupes à initiale occlusive ou fricative. Les données intégrales pour chaque groupe sont présentées en Annexe 10 (p.XV). Dans une seconde étape, nous étudions la distribution des fréquences individuelles de chaque groupe en fonction de sa catégorie phonétique. De même que dans l'analyse précédente, on observe des corrélations importantes entre les divers indices dérivés de cette analyse. Ainsi, le nombre d'occurrences des groupes de consonnes et la fréquence des mots porteurs sont très fortement corrélés (r = .97). Il est donc à nouveau possible d'affirmer que les groupes de consonnes fréquents en position initiale de mot ont tendance à apparaître dans des mots également fréquents.

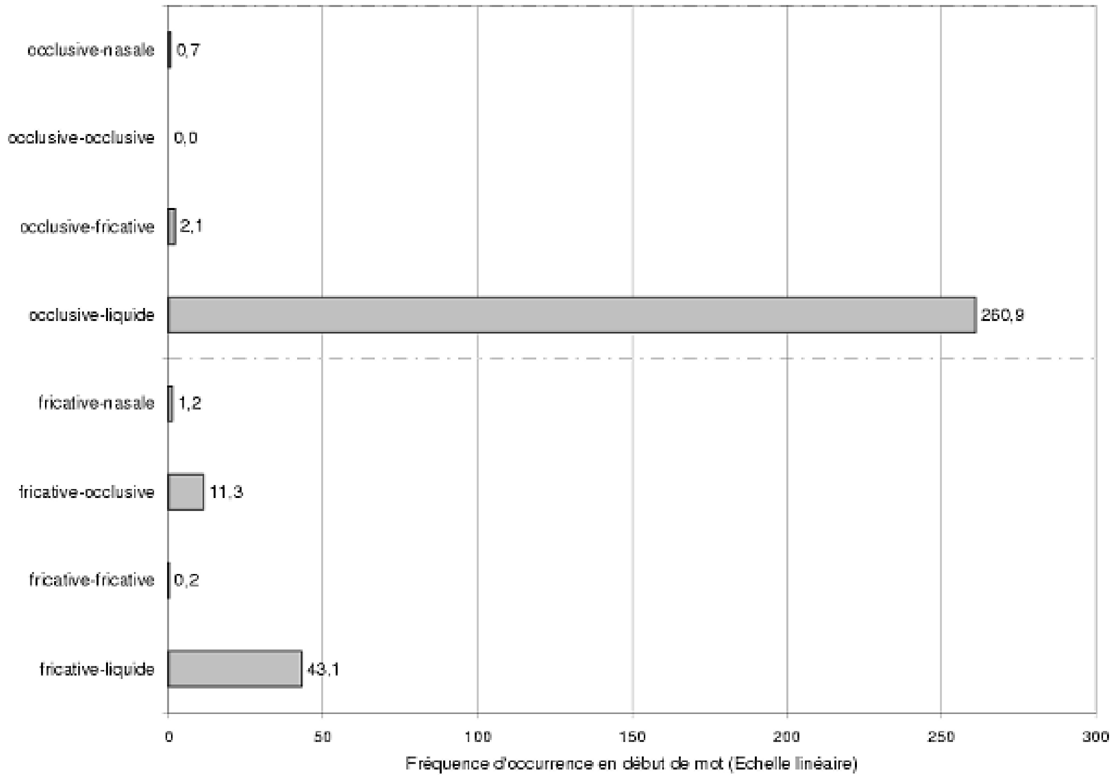

Figure 23 : Nombre moyen d'occurrences en début de mot. Groupes de consonnes classés en fonction de leur mode d'articulation.

3.2.1. Moyennes

La première étape de notre analyse consiste à étudier la fréquence moyenne (selon les 2 indices décrits précédemment) des groupes de consonnes en fonction de leurs caractéristiques phonétiques de mode d'articulation. Les graphiques suivants illustrent les différences de fréquence moyenne en position initiale de mot en fonction de la catégorie phonétique de mode d'articulation. Dans le premier (Figure 23) sont présentées les données moyennes obtenues pour le comptage du nombre de mots dans lesquels chaque groupe consonantique est recensé en position initiale. Nous ne présentons pas dans ce graphique les données obtenues pour les séquences CV (Consonne-Voyelle). Le second graphique reproduit les données de probabilité d'occurrence en début de mot (fréquence du groupe en position initiale pondérée par la fréquence d'usage des mots, Figure 24). Dans ce graphique nous présentons, outre les résultats observés pour les groupes de consonnes, les données obtenues pour les séquences Consonne-Voyelle.

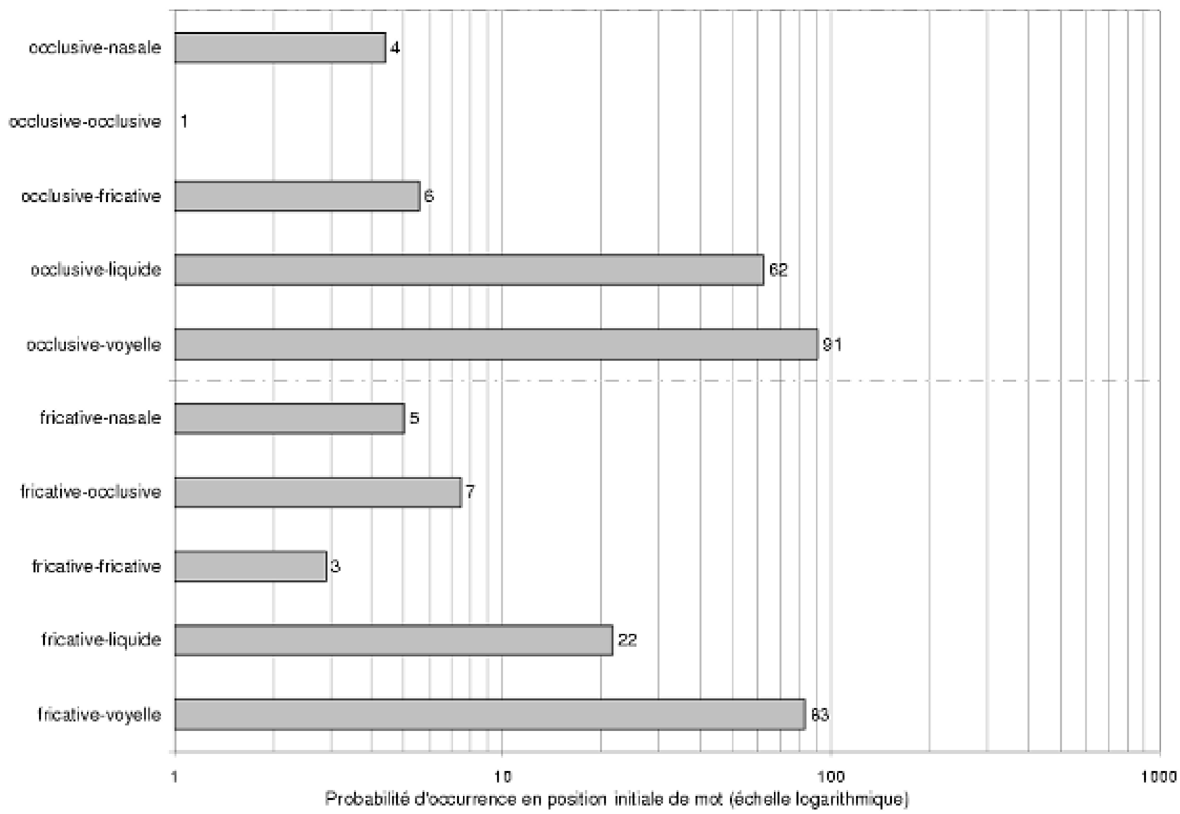

Figure 24 : Probabilité moyenne d'occurrence en début de mot. Groupes de consonnes classés en fonction de leur mode d'articulation.

On observe à nouveau une tendance des groupes tautosyllabiques à être plus fréquents que les groupes hétérosyllabiques. Les séquences C-liquide présentent une probabilité moyenne d'occurrence plus élevée (62 pour les groupes à initiale occlusive et 22 pour ceux à initiale fricative) que les autres catégories (respectivement 1, 4 et 6 pour les groupes à initiale occlusive et 3, 5, 7 pour les groupes à initiale fricative). Les séquences C-voyelle présentent quant à elles des probabilités d'occurrence plus importantes que l'ensemble des autres catégories (respectivement 91 et 83 pour les occlusive-voyelle et les fricative-voyelle). Les mêmes comparaisons *post-hoc* que dans l'analyse précédente ont été effectuées (test de Scheffé) en se restreignant aux groupes constitués d'une fricative ou d'une occlusive à l'initiale et en comparant, pour chaque ensemble, les données obtenues en fonction de la classe phonétique du second phonème. Les seuils de probabilité des diverses comparaisons effectuées à l'aide du test de Scheffé sont présentées dans le Tableau 5.

Cette analyse fournit des résultats tout à fait semblables à ceux qui ont été présentés dans l'étude des fréquences que nous avons conduite sans prendre en compte la position dans les mots. On observe notamment que, dans le cas des groupes à initiale occlusive, l'ensemble des comparaisons effectuées entre groupes *a priori* tautosyllabiques et hétérosyllabiques (l'intérieur du rectangle) permet de conclure à des différences significatives de fréquence. Il émerge en outre une différence significative entre séquences occlusive-voyelle et occlusive-liquide, différence qui avait déjà été observée dans les comparaisons de fréquence d'occurrence effectuées dans la première analyse. En ce qui concerne les groupes à initiale fricative, on observe également une certaine variabilité des seuils de probabilité obtenus. De même que dans la précédente analyse, ces groupes présentent une plus grande variabilité dans la significativité des différences relevées entre séquences tautosyllabiques et hétérosyllabiques. Les mêmes explications peuvent rendre compte de ce phénomène (variabilité plus importante entre structure syllabique déterminée a priori sur la base de leur mode d'articulation et structure syllabique effective en position intervocalique.

Tableau 5 : Seuils de probabilité des tests de Scheffé appliqués à la comparaison de probabilités d'occurrence des groupes de consonnes à initiale occlusive (a) ou fricative (b) apparaissant en position initiale de mot. Les comparaisons sont effectuées par catégorie de groupe en fonction du mode d'articulation des phonèmes. Les seuils de probabilité statistiquement significatifs sont retranscrits en caractères gras et italique. Les intitulés des lignes et des colonnes correspondent au second phonème de la séquence.

a/ Occlusive initiale	liquide	fricative	occlusive	nasale
voyelle	***0.000***	***0.000***	***0.000***	***0.000***
liquide		***0.000***	***0.000***	***0.000***
fricative			0.695	0.999
occlusive				0.935

b/ Fricative initiale	liquide	fricative	occlusive	nasale
voyelle	***0.010***	***0.000***	***0.000***	***0.000***
liquide		***0.016***	0.129	0.098
fricative			0.823	0.995
occlusive				0.990

Dans l'analyse précédente, nous avons vu qu'à ce lien entre fréquence et tautosyllabicité ne correspondait pas nécessairement une distinction nette en termes de fréquence individuelle des groupes de consonnes. Ainsi l'on observe, dans le cadre de l'étude des distributions de fréquence quelle que soit la position des groupes de consonnes dans les mots, des recouvrements

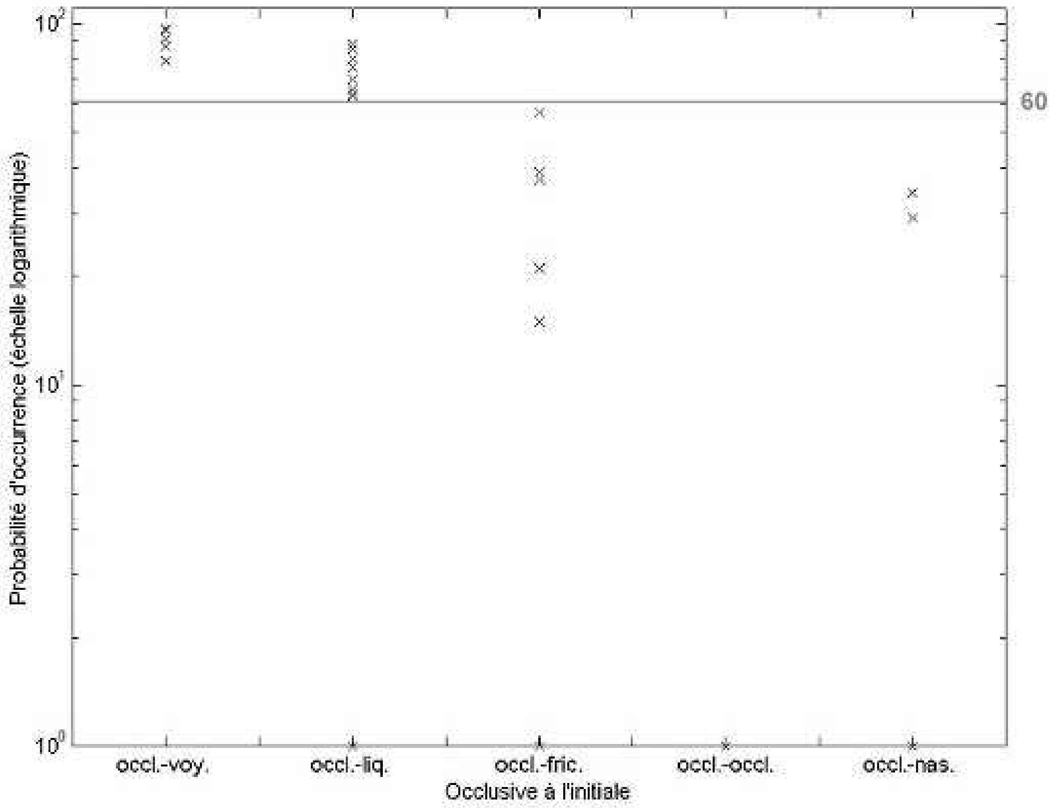

Figure 25 : Distribution des probabilités d'occurrence en début de mot pour les groupes de consonnes à initiale occlusive.

importants des distributions supposées désigner d'une part les groupes tautosyllabiques et, d'autre part, les groupes hétérosyllabiques. Le lien entre tautosyllabicité et fréquence n'étant pas discriminant, nous souhaitons obtenir un indice de bonne forme en position initiale de syllabe qui pourrait à la fois être observable dans un corpus de la langue et fournir une procédure de classification qui soit la plus discriminante possible. Il nous semble par conséquent intéressant d'étudier les distributions de probabilité d'occurrence en début de mot. Nous faisons l'hypothèse que, même si les données moyennes de probabilité d'occurrence fournissent des informations tout à fait semblables à celles qui ont été dérivées de la première analyse distributionnelle, la restriction aux débuts de mots pourrait fournir une illustration totalement différente de la distinction entre groupes tautosyllabiques et hétérosyllabiques si l'on s'intéresse aux distributions des valeurs mesurées. Notamment, nous supposons que cet indice devrait donner lieu à des recouvrements beaucoup moins importants que celui obtenu sans tenir compte de la position des groupes de consonnes dans les mots de la langue. Si c'est effectivement le cas, nous pourrions utiliser cet indice de fréquence en début de mot pour dissocier opérationnellement les groupes de consonnes en fonction de leurs caractéristiques phonologiques.

3.2.2. Distributions

Nous présentons ici les graphiques correspondant aux distributions de probabilité d'occurrence des groupes de consonnes en les classant par catégorie phonétique de mode d'articulation. La Figure 25 illustre les distributions des divers groupes de consonnes commençant par une occlusive (/b/, /d/, /g/, /p/, /t/, /k/). Les données correspondant aux groupes à initiale fricative (/v/, /z/, /ʒ/, /f/, /s/, /ʃ/) sont présentées dans la Figure 26.

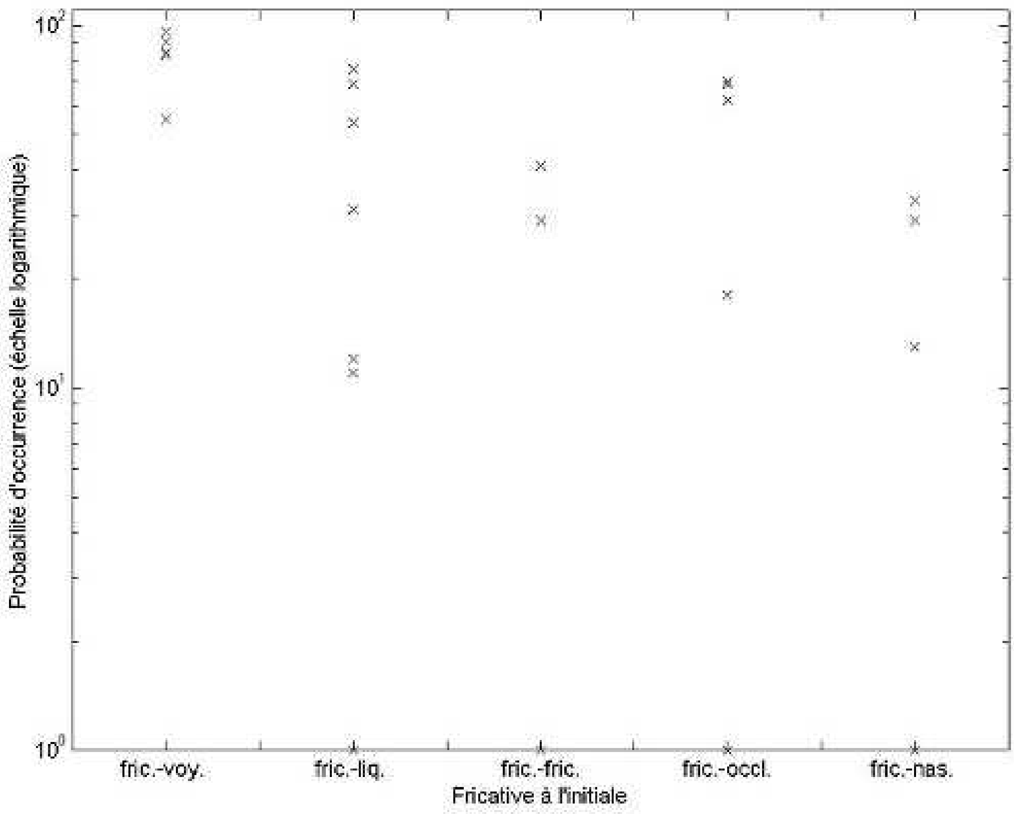

Figure 26 : Distribution des probabilités d'occurrence en position initiale de mot pour les groupes de consonnes à initiale fricative.

Dans le graphique correspondant aux groupes à initiale occlusive, on observe des distributions de probabilité d'occurrence tout à fait différentes de celles obtenues dans la précédente analyse. Contrairement à la distribution des valeurs obtenues sans tenir compte de la position des groupes dans les mots, la restriction du calcul de la fréquence aux groupes apparaissant en position initiale de mot fait émerger une distinction nette entre groupes tautosyllabiques et hétérosyllabiques. Nous avons tracé, dans la Figure 25, une droite horizontale indiquant cette frontière (la valeur discriminante étant approximativement à 60). Les séquences occlusive-voyelle et occlusive-liquide ont presque toutes une probabilité d'occurrence supérieure à cette valeur (à l'exception des occlusive-liquide coronales /tl/ et /dl/) alors que toutes les autres séquences ont sans exception une valeur de probabilité d'occurrence inférieure à cette limite.

L'analyse des distributions de probabilité d'occurrence des groupes à initiale fricative est nettement plus délicate. On observe des distributions très similaires à celles qui ont été obtenues dans l'analyse globale des probabilités d'occurrence. Cette similarité, de même que la forme très différente des distributions en position initiale de mot selon que l'on s'intéresse aux groupes à initiale occlusive ou fricative, nous conduit à réitérer la remarque d'une variabilité beaucoup plus importante des structures syllabiques possibles dans la catégorie des groupes à initiale fricative. Cette observation est certainement liée à des paramètres phonétiques ou phonologiques. En l'occurrence, il est probable que la classification choisie en termes de mode d'articulation n'est pas la mieux appropriée à une distinction entre groupes tautosyllabiques et hétérosyllabiques.

3.3. Discussion

Que l'on s'intéresse aux probabilités d'occurrence dans la langue quelle que soit la position des séquences dans les mots ou au contraire à celles qui peuvent être mesurées en position initiale de mot, on observe un lien non négligeable entre la fréquence des suites de phonèmes et le statut d'*attaque syllabique bien formée*. Ce lien se manifeste par des différences de fréquence importantes entre diverses catégories de groupes consonantiques catégorisées sur la base du mode d'articulation des phonèmes qui les constituent. Nous avons néanmoins observé que ce lien pouvait n'être pas très stable. Notamment, les données obtenues pour les groupes à initiale fricative ne permettent pas de dissocier clairement, sur la base de la fréquence, les diverses catégories de groupes comparées dans notre analyse. On peut cependant légitimement considérer que la classification adoptée *a priori* n'est pas optimale pour séparer, parmi l'ensemble des groupes à initiale fricative, ceux qui constituent des attaques de syllabe bien formées et les autres, lesquels seraient alors nécessairement hétérosyllabiques en position intervocalique. Il est probable qu'une classification mieux adaptée aurait conduit à des résultats plus convaincants. L'analyse des données concernant les groupes à initiale occlusive est beaucoup plus pertinente. Il semble que la classification adoptée (à l'exception des deux groupes occlusive-liquide coronaux /tl/ et /dl/) soit bien adaptée à la distinction entre groupes tautosyllabiques et groupes hétérosyllabiques parmi les séquences à initiale occlusive. On observe notamment que cette classification permet de dissocier assez clairement deux catégories de groupes de consonnes à l'aide de tests statistiques. Les diverses comparaisons effectuées à l'aide du test de Scheffé permettent en effet de distinguer très clairement les séquences C-voyelle et C-liquide des 3 autres catégories. Or nous avons supposé dès le départ que ces deux premières catégories devraient en général constituer des attaques syllabiques bien formées. Les 3 autres étaient à notre avis constituées de groupes ne constituant pas réellement des attaques syllabiques légales. La plupart des éléments de ces catégories pouvaient en réalité correspondre à ce que Dell (1995) désigne comme groupes légaux mais déviants. Les données dérivées de l'analyse des groupes à initiale occlusive fournissent donc un argument de poids pour affirmer l'existence d'un lien entre la légalité d'une séquence en attaque de syllabe et sa fréquence dans la langue.

Si ces deux paramètres sont intimement liés, ils ne sont pas confondus. C'est ce qu'illustre l'étude des distributions de probabilité d'occurrence. Ces distributions présentent des recouvrements importants qui mettent en évidence une certaine dissociation entre fréquence et légalité. De fait, nous avons choisi de rechercher un indice qui permettrait de dissocier clairement deux ensembles de groupes de consonnes sur la base de leur légalité en attaque syllabique et qui serait observable dans la langue. L'analyse distributionnelle conduite sur les débuts de mots permet de discriminer nettement les catégories de groupes de consonnes à partir d'une information statistique. Alors que l'étude des distributions de probabilités brutes ne permet pas de dissocier tautosyllabicité et fréquence, l'étude des distributions de probabilité d'occurrence en début de mot fait émerger pour les groupes qui sont attestés dans l'échantillon une discrimination très nette entre d'une part les groupes occlusive-liquide qui sont tous fréquents en début de mot et les 3 autres catégories (occlusive-fricative, occlusive-occlusive et occlusive-nasale) qui sont tous nettement plus rares. L'absence de recouvrement entre les distributions est un indicateur particulièrement utile du lien avec les contraintes phonologiques de la langue. Ainsi, alors que la fréquence brute peut être assimilée à un observable corrélatif mais dissociable des régularités phonologiques, la fréquence d'occurrence en début de mot refléterait fidèlement ces régularités et constituerait une mesure pertinente de la distinction entre séquences légales et illégales en attaque syllabique.

4. Réanalyse des données comportementales

L'analyse distributionnelle conduite sur la base de données BRULEX (Content et al., 1990) fournit les fondements d'une critique raisonnée des données comportementales avancées comme des preuves du recours, de la part des locuteurs natifs, à des processus de segmentation de la parole en mots qui reposeraient en partie sur des connaissances concernant les régularités phonologiques de leur langue. Trois ensembles d'interprétations peuvent en réalité être proposées. Nous procédons en premier lieu à un rapide rappel des données obtenues. Après avoir décrit à nouveau l'interprétation favorisée par ces auteurs, nous proposons 3 interprétations permettant de prédire ces effets. Aucune d'entre elles ne nécessite le recours à des processus faisant intervenir des connaissances sur les régularités de la langue.

4.1. Rappel des données

Les données expérimentales présentées par (McQueen, 1998) et par Vroomen & De Gelder (1999) ont été mises en évidence à l'aide de deux paradigmes expérimentaux différents : le *word-spotting* et la détection de phonèmes. Les tâches ayant déjà été décrites dans le Chapitre 2 (Sections 2.2.3.2 et 2.2.3.3), nous passons immédiatement au rappel des diverses conditions expérimentales comparées et aux interprétations fournies par les auteurs respectifs de ces études.

4.1.1. Word-spotting

McQueen (1998) a étudié le rôle des contraintes phonotactiques dans les processus de segmentation de la parole en mots avec des locuteurs de langue maternelle néerlandaise. Il reprend la tâche de *word-spotting* introduite par Cutler & Norris (1988) en manipulant le statut du groupe de consonnes médian. Les mots à détecter apparaissent soit en position initiale soit en position finale du non-mot. La légalité phonotactique du groupe consonantique médian est manipulée, celui-ci étant phonotactiquement légal (/vr/, /dr/) ou illégal (*/mr/, */nr/). Lorsque le mot à détecter est en position initiale, le groupe légal donne lieu à un non-alignement de la frontière phonotactique avec la frontière lexicale. Si le mot à détecter est *pil* ('pilule'), le stimulus /pɪlvrem/ -qui contient le groupe consonantique médian légal /vr/- donne lieu à une correspondance entre segmentation phonotactique /pil.vrem/ et segmentation lexicale /pil.vrem/. Lorsque le groupe médian est illégal (par exemple */mr/ dans /pɪlmrem/), il n'y a par contre plus correspondance entre segmentations phonotactique /pilm.rem/ et lexicale /pil.mrem/. Au contraire, si le mot à détecter est en position finale, on observe une relation inverse entre légalité du groupe médian et alignement des frontières. Pour une détection du mot *rok* ('jupe'), la séquence /fimrɒk/ (dans laquelle */mr/ est illégal) donne lieu à un alignement de ces frontières -avec une segmentation phonotactique /fim.rɒk/- alors qu'un groupe de consonnes légal (par exemple /dr/ dans /fidrɒk/) induit un non-alignement des découpages phonotactique (/fi.drɒk/) et lexical (/fid.rɒk/).

Tableau 6 : Statut du groupe de consonnes médian dans l'expérience de McQueen (1998) en fonction de l'alignement entre frontières phonotactique et lexicale.

	non-alignement	**alignement**
Position initiale	illégal /pɪlmrem/	légal /pɪlvrem/
Position finale	légal /fidrɒk/	illégal /fimrɒk/

Le Tableau 6 reprend les informations présentées dans le Tableau 2 du Chapitre 2 et résume l'agencement du lien entre légalité et alignement dans cette expérience. McQueen (1998) observe qu'une absence d'alignement entre les frontières phonotactique et lexicale donne lieu à des taux d'erreur significativement plus importants que la condition d'alignement. Les mots sont détectés plus facilement lorsque frontière phonotactique et lexicale sont alignées que lorsqu'elles sont discordantes.

4.1.2. Détection de phonèmes

Vroomen & De Gelder (1999) présentent à des locuteurs néerlandais des phrases dans lesquelles les participants doivent détecter un phonème-cible. Selon les conditions, le phonème-cible peut être prononcé en fin de syllabe comme dans la séquence :

'de.boo**t**.die.ge.zon.ke.nis'

dans laquelle la cible est le phonème /t/. Dans cette situation, on observe qu'il y a correspondance entre la frontière syllabique qui sépare /but/ et /di/ et la frontière lexicale qui sépare *boot* ('bateau') et *die* ('qui'). Dans l'autre condition expérimentale, le phonème-cible apparaît en position d'attaque syllabique comme dans l'énoncé :

'de.boo.**t**is.ge.zon.ken'

On observe ici une discordance entre la frontière syllabique (laquelle sépare /bu/ de /tiz/) et la frontière lexicale qui se situe après *boot* ('bateau'). Dans la première condition, la consonne cible (une occlusive) est suivie d'une consonne également occlusive qui induit un alignement des frontières syllabique et lexicale. Dans la condition de non-alignement des frontières, cette consonne cible est au contraire suivie d'une voyelle. Les auteurs observent également un effet de la relation entre frontière syllabique et frontière lexicale. Cet effet se manifeste cependant ici dans les temps de réaction mais pas dans les taux d'erreur : les latences de détection de phonème sont plus courtes lorsque les frontières syllabique et lexicale sont alignées (c'est à dire lorsque le phonème-cible est en position de coda syllabique) que lorsqu'elles ne le sont pas (phonème en attaque syllabique).

4.1.3. Interprétations

Les auteurs aboutissent à des conclusions similaires à l'issue de l'analyse des données expérimentales. Dans les deux tâches, l'identification du mot pertinent serait influencée par le découpage syllabique ou phonotactique de la chaîne de phonèmes. Lorsque ce découpage concorde avec la segmentation lexicale adéquate, les processus de reconnaissance lexicale seraient facilités, ceci se manifestant par un raccourcissement des délais de réalisation de la tâche ou par un accroissement des taux de réponse correcte. De fait, ces données constitueraient une mise en évidence du recours à des connaissances sur les régularités phonologiques de la langue dans les processus de segmentation lexicale qui sont mis en œuvre au cours du traitement du signal de parole. Nous avons cependant mis en évidence un lien entre tendance des séquences de phonèmes à se regrouper à l'attaque syllabique et fréquence d'occurrence dans la langue.

4.2. Interprétations concurrentes

Il nous semble possible, sur la base du lien observé entre tautosyllabicité et fréquence, de proposer au moins 3 classes d'interprétations de ces effets. Chacune de ces trois interprétations est directement dérivée de la correspondance observée entre légalité phonologique des séquences de phonèmes et fréquence d'occurrence de ces mêmes suites dans les mots de la langue. Nous avons montré dans l'analyse distributionnelle qui a été conduite sur la base de données BRULEX (Content et al., 1990) que, dans un lexique français, les groupes légaux sont utilisés moins fréquemment dans les mots de la langue que les groupes illégaux. Il est probable, si ce lien est effectivement une conséquence des phénomènes de régularité phonologique, que les mêmes résultats seraient obtenus en néerlandais. Dans l'expérience de McQueen (1998), les groupes identifiés comme légaux sont des séquences occlusive-liquide ou fricative liquide (parmi lesquelles /dr/, /vr/, /fl/…) alors que les groupes illégaux sont des séquences nasale-liquide (les deux suites /mr/ et /nr/) ou la suite /tl/. La même classification pourrait en fait être adoptée en français. Ainsi, les suites illégales utilisées par McQueen (1998) n'apparaissent jamais au début des mots de la langue française alors que les suites occlusive-liquide non-coronales et fricative-liquide sont courantes dans cette position. Or nous avons mis en évidence des différences de fréquence considérables entre ces deux catégories de groupes. La même critique peut s'appliquer aux données de Vroomen & De Gelder (1999). En français, les séquences CV sont nettement plus fréquentes que les suites de consonnes hétérosyllabiques. Il semble donc y avoir une confusion, dans les variables manipulées par ces auteurs, entre légalité / tautosyllabicité et fréquence d'occurrence dans la langue. Cette observation nous conduit à proposer trois interprétations alternatives des données expérimentales présentées.

4.2.1. Un phénomène de sélection lexicale sérielle ?

La première interprétation que nous proposons est inspirée du modèle COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987) de sélection lexicale que nous avons décrit dans le Chapitre 2. Selon ce modèle, le processus de reconnaissance des mots dans la modalité auditive génère l'activation d'un ensemble de candidats à partir des premières périodes acoustiques du stimulus. Les informations acoustiques ultérieures donnent lieu à une sélection lexicale consistant à supprimer de l'ensemble initial de mots activés (*la cohorte*) les candidats qui ne sont plus appariés avec l'entrée acoustique. Nous avons décrit la possibilité introduite par ce modèle de reconnaître un mot avant même sa fin acoustique. Nous avons également cité plusieurs études computationnelles et comportementales qui mettent en évidence la possibilité qu'un mot soit reconnu après sa fin acoustique s'il existe dans le lexique d'autres mots au début desquels le mot à identifier est enchâssé (Luce, 1986; Frauenfelder & Peeters, 1990; Grosjean, 1985). Ces deux phénomènes sont à mettre en rapport avec les données auxquelles nous nous intéressons ici. Dans les expériences qui ont permis de mettre en évidence un rôle des régularités phonologiques dans les processus de segmentation lexicale, on a contrôlé les caractéristiques des groupes de consonnes comparés en fonction de leur regroupement éventuel à l'attaque de syllabe. Cette distinction entre séquences de phonèmes regroupées à l'attaque syllabique et séquences nécessitant l'insertion d'une frontière phonologique entre les deux phonèmes correspond dans la langue à une distinction en termes de fréquence d'occurrence. Les groupes de phonèmes qui se regroupent à l'attaque de syllabe sont plus fréquemment utilisés dans les mots de la langue. Or, si les séquences légales sont plus souvent utilisées dans les mots de la langue, il est légitime de prédire qu'une séquence CVC donnée aura plus de chances de constituer le début d'un mot pour lequel la consonne finale de la séquence CVC et le phonème suivant (consonne ou voyelle) constitueront une suite fréquente (donc une attaque de syllabe bien formée) qu'une suite rare (illégale ou hétérosyllabique). Ainsi, le monosyllabe 'vague' constitue le début de 14 mots dans la langue. Pour 2 d'entre eux la consonne /m/ fait suite à la séquence /vag/, créant ainsi une séquence hétérosyllabique /gm/. Par contre, les 12 autres mots font suivre /vag/ par une voyelle, induisant alors la consonne /g/ à se retrouver à l'attaque de la syllabe suivante. Une recherche rapide dans la base de données BRULEX (Content et al., 1990) conduit à cette même observation pour nombre de séquences de 3 phonèmes ayant cette structure CVC. Certaines suites donnent lieu à un déséquilibre encore plus important entre les rattachements possibles de la consonne finale. Une recherche des mots possibles commençant par /bat/ fait ressortir 36 mots au début desquels cette séquence est enchâssée. Trois d'entre eux font suivre le /t/ d'un /l/ ou d'un /m/, donnant ainsi lieu à une séquence hétérosyllabique. L'ensemble des autres mots possibles fait suivre le /t/ d'un /r/ ou d'une voyelle. Lorsque l'on traite une suite CVC dans une tâche d'identification de mot, il est par conséquent très probable que le phonème suivant sera regroupé à l'attaque de syllabe avec la consonne finale de la séquence CVC. Cette probabilité est valable aussi bien dans une tâche de détection de phonème (qui induit souvent le recours à l'identification du mot porteur) que dans une tâche de *word-spotting* (qui l'induit nécessairement). Supposons par exemple que l'on souhaite conduire une expérience avec le paradigme de *word-spotting*, expérience dans laquelle on chercherait à mettre en évidence le rôle des régularités phonologiques dans les processus de segmentation de la parole en mots. L'une des cibles que les participants devront détecter est le mot 'bague'. Dans la condition d'alignement entre segmentation phonologique et segmentation lexicale, la consonne /g/ est suivie d'un /n/. Dans l'autre condition expérimentale, on fait suivre cette même consonne du phonème /a/. Lorsque l'on a traité la suite /vag/, il reste dans le lexique un ensemble de candidats lexicaux constitués du mot 'bague' en position initiale. Ces mots sont au nombre de 14 dans la base de données BRULEX (Content et al., 1990). Il n'est donc pas possible d'identifier le mot 'bague' à partir de la sélection des candidats lexicaux présents dans la cohorte puisqu'ils sont trop nombreux pour cela. Huit d'entre eux commencent par /baga/ alors qu'un seul commence par /bagn/. Une fois le phonème suivant traité et utilisé pour procéder à la suppression des candidats qui ne sont plus appariés avec l'entrée sensorielle, la quantité de candidats maintenus dans la cohorte en fonction de la condition expérimentale subit un déséquilibre. Dans la condition de légalité phonotactique-tautosyllabicité, il faudra sélectionner le mot adéquat 'bague' parmi une quantité plus importante de candidats (Les huit mots commençant par /baga/ plus le mot 'bague') que dans la condition d'illégalité phonotactique-hétérosyllabicité dans laquelle seul le mot 'baguenauder' sera maintenu dans la cohorte. Bien évidemment, nous avons spécifiquement choisi cet exemple afin de mettre en évidence une possibilité de déséquilibre lexical dans la constitution d'un matériel destiné à mettre en évidence des effets non-lexicaux. Il est cependant nécessaire d'admettre que si les phonèmes qui font suite à la consonne finale du mot cible (dans la tâche de *word-spotting*) ou du mot porteur (dans la tâche de détection de phonèmes) sont choisis au hasard, la probabilité de choisir des phonèmes donnant lieu à une quantité plus importante de candidats lexicaux dans la condition de légalité phonotactique que dans la condition d'illégalité phonotactique est grande. A moins d'être averti de ce risque et de contrôler ce paramètre, on risque d'observer des effets lexicaux qui pourraient passer pour des effets phonologiques.

4.2.2. Segmentation probabiliste

Plusieurs études récentes ont conduit à affirmer que le système de traitement du langage serait en mesure d'utiliser des informations probabilistes pour traiter les informations de l'environnement linguistique (Brent, 1996 ; Brent & Cartwright, 1996). Ces procédures statistiques reposeraient sur des mécanismes généraux de traitement (Aslin, Saffran, & Newport, 1998; Saffran, Johnson, Aslin, & Newport, 1999) mais pourraient être appliquées à des processus dédiés au traitement du langage. Il est possible d'observer le recours à des informations probabilistes dès les premiers mois (Saffran, Aslin et al., 1996). Dans le cadre de l'acquisition du langage, ces mécanismes auraient pour fonction de déclencher la constitution d'un lexique initial qui permettrait par la suite de développer des processus de reconnaissance des mots fondés par exemple sur les compétitions entre candidats lexicaux (McClelland & Elman, 1986). A l'âge adulte, ils continueraient d'être utilisés et permettraient de prédire les frontières entre les mots (Saffran, Newport et al., 1996 ; Brent, 1997). Toute séquence de phonèmes rencontrée dans la chaîne de parole serait codée en termes de probabilité transitionnelle d'apparition. Une séquence très fréquente conduirait le système à supposer que cette séquence fait partie d'un mot unique. Il aurait alors tendance à la regrouper afin de chercher dans son lexique des candidats contenant cette séquence. Une séquence très rare ayant peu de chances d'exister effectivement dans un mot de la langue, les phonèmes la constituant seraient au contraire considérés comme faisant partie de deux mots différents et le système aurait tendance à chercher des séquences de mots en insérant une frontière lexicale entre les phonèmes de la séquence rare. Du fait du lien que nous avons mis en évidence entre fréquence et légalité, il est en réalité difficile d'affirmer que les données avancées comme des preuves du recours à des connaissances sur les régularités de la langue ne constituent pas au contraire un reflet de l'utilisation de calculs probabilistes consistant à séparer les séquences de phonèmes. Des travaux portant sur le traitement langagier chez le jeune enfant et le nourrisson ont conduit à affirmer que ces deux catégories d'informations sont disponibles pour les processus de traitement de la parole. Dès les premiers mois de la vie, les enfants auraient intégré des connaissances concernant la légalité phonotactique des séquences de phonèmes dans leur langue maternelle (Jusczyk, Luce, & Charles-Luce, 1994; Friederici & Wessels, 1993). Ils présentent par ailleurs une tendance à considérer comme plus familières des séquences fréquentes que des séquences rares dans leur langue (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). En fait, il n'est pas sûr que ces deux catégories de processus ne constituent pas deux facettes d'un même phénomène. En effet, puisque fréquence et légalité phonotactique sont liées, il est possible que les comportements observés chez l'enfant lorsqu'il entend des suites de phonèmes phonotactiquement illégales soient déterminés par des différences de fréquence plus que par des différences de légalité phonotactique impliquant le recours à une base de connaissances linguistiques. A l'inverse, on peut tout aussi bien faire l'hypothèse que les effets probabilistes observés consistant à trouver plus familière une séquence fréquente dans la langue pourraient être déterminés par des représentations linguistiques qui permettraient à l'enfant de considérer des séquences fréquentes comme *plus légales* que des séquences rares.

Nous sommes confrontés au même dilemme dans l'étude des processus de segmentation lexicale chez l'adulte. Si l'on met en évidence un effet de la légalité phonotactique ou de la structure syllabique sur les temps de réaction observés, on peut légitimement se demander si les effets interprétés en termes phonologiques ne sont pas tout simplement déterminés par des représentations de type probabiliste. Le système de traitement de la parole pourrait ainsi localiser des groupes de phonèmes rares et faire l'hypothèse d'une frontière lexicale entre les phonèmes constituant des séquences rares dans la langue. Il ne serait donc pas nécessaire, même si les processus semblent très similaires, d'avoir recours à un modèle dans lequel les locuteurs feraient appel à des connaissances sur les régularités linguistiques de leur langue.

4.2.3. Fréquence et compétitions lexicales

Les deux interprétations précédentes font appel à deux types de processus différents. Selon la première, la tendance à répondre plus lentement lorsque l'on entend une séquence pour laquelle il existe un non-alignement entre frontière phonotactique / syllabique et frontière lexicale pourrait s'expliquer par le maintien d'une quantité plus importante de candidats lexicaux dans la cohorte (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978) lorsque le phonème qui suit la consonne finale du mot est en cours de traitement. Ce phénomène implique un processus de traitement séquentiel consistant à sélectionner progressivement les candidats les mieux appariés avec l'entrée acoustique. Cette interprétation suppose que l'effet observé serait localisé à un niveau lexical de traitement. Une seconde interprétation alternative des données obtenues consiste à proposer que cette différence de statut des séquences de phonèmes manipulées dans les expériences impliquerait une segmentation fondée sur des différences de fréquence des séquences plus que sur des différences de statut phonologique. L'effet observé serait alors également interprétable en termes pré-lexicaux : la présence d'une séquence rare (de même que celle d'une séquence illégale ou hétérosyllabique) conduirait à supposer une frontière lexicale entre les phonèmes qui la constituent.

Une troisième interprétation nous semble envisageable. Dans le cadre d'un modèle de compétitions lexicales comme TRACE (McClelland & Elman, 1986) ou SHORTLIST (Norris, 1994), les candidats lexicaux activés n'ont pas besoin d'être alignés avec le début du mot comme c'est le cas dans COHORT (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). On parle d'alignement exhaustif des activations lexicales. Dans ces modèles, des processus d'activation interactive entre les divers candidats activés sont implémentés (cf. Chapitre 2). Le traitement effectué sur l'entrée acoustique n'est donc pas intégralement sériel. Il n'est par conséquent pas totalement nécessaire de coder les phonèmes les uns après les autres et de sélectionner progressivement le candidat adéquat en respectant l'axe temporel de prononciation des phonèmes. L'interprétation lexicale proposée dans la section 4.2.1 ne serait donc pas envisageable avec cette classe de modèles. Par contre, si une séquence de phonèmes est fréquente ceci signifie qu'elle apparaît dans une quantité importante de mots de la langue. L'occurrence d'une séquence fréquente conduirait donc à provoquer l'activation d'un grand nombre de candidats lexicaux, quelle que soit la position de la séquence dans les mots. La procédure de compétition entre les candidats activés impliquerait par conséquent une quantité d'unités lexicales beaucoup plus importante que dans le cas de l'occurrence d'une séquence de

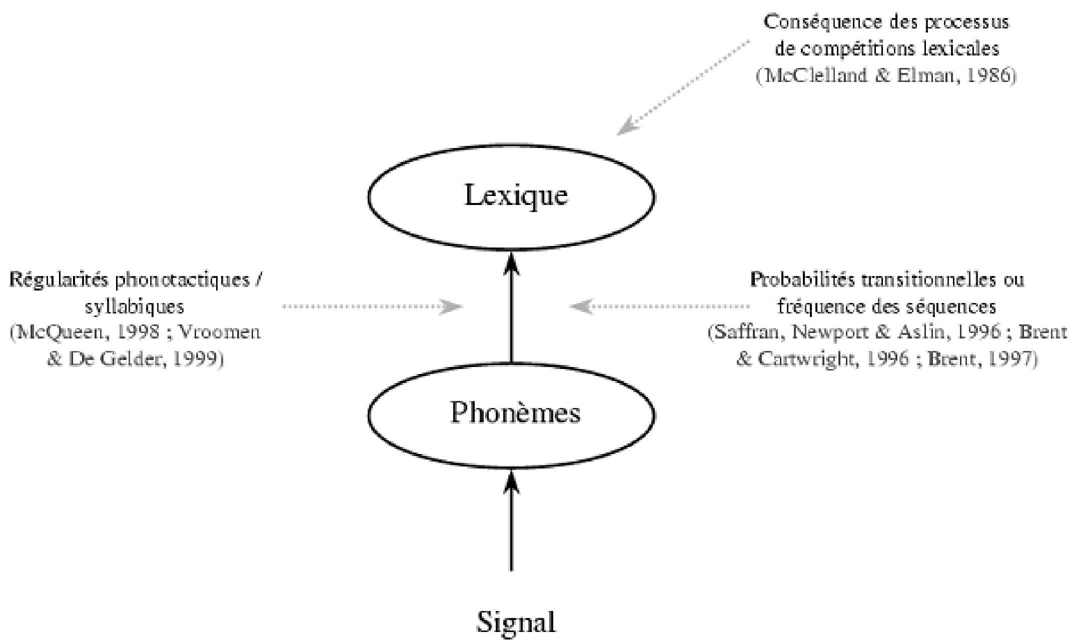

Figure 27 : Représentation graphique des interprétations reposant sur la fréquence des groupes de consonnes (effet prélexical pour une segmentation probabiliste ; effet lexical pour les phénomènes de compétitions lexicales.

phonèmes rare. Il faudrait alors plus de temps au réseau pour aboutir à une stabilisation des niveaux d'activation et, de fait, à la sélection du candidat approprié.

Cette dernière interprétation doit déjà être mise à l'épreuve des données qui sont présentées par McQueen (1998). En effet, lorsqu'il étudie l'effet de l'alignement entre frontières phonotactique et lexicale avec le mot en position initiale, la condition d'alignement (la plus facile) correspond à une séquence phonotactiquement légale (qui, en raison du nombre de mots de la langue dans lesquels elle apparaît, générerait l'activation d'une importante quantité de candidats lexicaux). Or, si notre interprétation était valide, on devrait observer l'effet inverse puisque la séquence légale devrait alors rendre la tâche plus difficile et donner lieu à des taux d'erreur plus importants. La même remarque peut-être faite pour l'interprétation en termes de sélection lexicale sérielle. On notera néanmoins que les effets obtenus par McQueen (1998) ne semblent pas totalement liées au contexte phonotactique mais pourraient s'expliquer par des caractéristiques acoustiques des stimuli ; lesquels seraient prononcés différemment en contexte légal et illégal. Dans une seconde expérience, McQueen (1998) a extrait des stimuli utilisés dans la tâche de *word-spotting* le segment acoustique correspondant au mot et a conduit une tâche de décision lexicale avec ce matériel en étudiant à nouveau l'effet de l'alignement. Si l'effet original est réellement lié au contexte phonotactique, il devrait alors disparaître puisqu'aucun contexte n'est adjoint aux stimuli expérimentaux. C'est effectivement le cas pour les mots qui ont été prononcés en position finale de non-mot. L'effet d'alignement disparaît lorsqu'on réalise une tâche de décision lexicale sur la partie qui correspond au mot. Par contre, l'effet d'alignement se maintient avec les mots qui ont été prononcés à l'origine en position initiale. Il est donc probable que l'effet d'alignement obtenu dans la tâche de *word-spotting* était dans cette situation la conséquence de différences propres aux caractéristiques intrinsèques de la partie lexicale et n'était pas lié au contexte phonotactique. McQueen (1998) conduit une analyse de covariance (ANCOVA) sur les données de *word-spotting* avec comme covariable les taux d'erreurs obtenus dans la tâche de décision lexicale et observe que l'effet d'alignement observé en *word-spotting* se maintient. Il en conclut logiquement que l'effet observé dans la tâche de décision lexicale ne suffit pas à expliquer intégralement l'effet obtenu en *word-spotting*. L'interprétation initiale serait donc quand même valide. En réalité, la tâche de *word-spotting* est beaucoup plus difficile à effectuer que celle de décision lexicale. Lorsque l'on conduit une expérience avec la tâche de *word-spotting*, on observe des taux d'erreurs et des temps de réaction moyens beaucoup plus importants qu'en décision lexicale. Les participants fournissent aussi des appréciations sur la tâche qui montrent qu'elle est particulièrement difficile. Il suffit d'observer les taux d'erreurs moyens produits par les participants de l'expérience de McQueen (1998) qui vont de 20% (alignement) à 60% (non-alignement) pour s'en persuader. Dans une tâche de décision lexicale, on admettrait difficilement un taux d'erreur supérieur à 10% ! Il nous semble illégitime d'introduire comme covariable d'une ANCOVA des données similaires (taux d'erreur) ayant été obtenues avec une tâche qui s'avère beaucoup plus facile. En effet, si l'on compare les mêmes effets dans une tâche très facile et dans une tâche relativement difficile, on peut s'attendre à ce que les effets observés dans la première soient plus faibles que dans la seconde. On peut également s'attendre à obtenir une variabilité beaucoup plus importante dans la tâche difficile que dans la tâche facile. Or une ANCOVA consiste à évaluer la part de variance d'une variable que ne peut pas expliquer une seconde variable (la covariable). Si la non-correspondance des données est liée à une différence dans la difficulté intrinsèque des tâches et pas à l'intervention de deux catégories de facteurs dont l'un n'intervient pas dans la seconde tâche, il est impossible d'utiliser les données de l'ANCOVA pour tenter d'interpréter les données obtenues dans la première tâche.

Les données observées pour les mots en position initiale ne sont donc pas assez fiables pour permettre de réfuter ces hypothèses lexicales et il est nécessaire de répliquer les expériences afin de confronter les interprétations proposées par McQueen (1998) et par Vroomen & De Gelder (1999) à des données expérimentales permettant de contrôler les diverses variables qui peuvent intervenir dans l'émergence de ces effets.

Résumé

Une analyse distributionnelle a été conduite sur la base de données lexicales informatisée de la langue française BRULEX (Content et al., 1990). Cette analyse permet de mettre en évidence un lien étroit entre légalité phonotactique et fréquence d'occurrence des séquences de phonèmes dans la langue. Cette observation nous a conduit à réanalyser les données comportementales présentées par McQueen (1998) et par Vroomen & De Gelder (1999) en mettant en évidence une confusion entre le statut phonologique des types de séquences comparés et la fréquence des séquences dans la langue. Ce lien entre structure phonologique des groupes de consonnes et fréquence d'occurrence conduit à proposer 3 interprétations différentes des données expérimentales. Aucune de ces interprétations ne nécessite d'avoir recours à des connaissances sur la structure phonologique de la langue. Il est donc nécessaire d'approfondir l'étude du rôle des régularités phonologiques de la langue dans les processus de segmentation de la parole en mots afin

d'approfondir la compréhension de ces effets et des processus qui les sous-tendent.

Chapitre 5

Contraintes phonologiques et sélection lexicale

Ce chapitre a fait l'objet d'une communications orale dans un congrès avec comité de lecture assorti d'une publication dans des actes :

XXII^{èmes} Journées d'Etude sur la Parole, 15-19 Juin 1998, Martigny, Suisse.

CONTRAINTES PHONOLOGIQUES ET SELECTION LEXICALE

Dans cette première série d'expériences, nous confrontons les données expérimentales interprétées comme la preuve d'un recours à des connaissances sur la structure phonologique de la langue (McQueen, 1998; Vroomen & de Gelder, 1999) à la première interprétation alternative que nous avons proposée. Nous étudions par conséquent le rôle possible des phénomènes de sélection lexicale dans un modèle purement séquentiel de l'accès au lexique : le modèle COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987). Avant de passer plus précisément à l'analyse des rôles dissociés de la fréquence et des régularités phonologiques, il est en effet important de déterminer dans quelle mesure l'un des modèles qui a exercé le plus d'influence au début des années 80 peut rendre compte des résultats obtenus. Nous avons montré dans le chapitre précédent que fréquence et tautosyllabicité sont liées. Ceci implique qu'un choix aléatoire des séquences de phonèmes supposées permettre la comparaison de séquences phonotactiquement légales ou illégales risque fort de conduire à un déséquilibre dans la structure des stimuli utilisés. Le phonème donnant lieu à une séquence légale risque en effet de donner lieu à une séquence constituant le début d'un nombre de mots plus élevé que celui qui donne lieu à une séquence illégale. Lorsque ce phonème est en cours de traitement, le nombre de candidats lexicaux qui sont maintenus dans la cohorte peut différer en fonction du type de groupe de phonèmes comparé. Il est donc nécessaire de contrôler cette variable afin de restreindre le champ des interprétations possibles.

1. Expérience 1 : Contraintes phonologiques et sélection lexicale

Dans cette première expérience, nous étudions l'effet de la structure phonologique des groupes de consonnes en contrôlant l'influence éventuelle de phénomènes de sélection lexicale sérielle en nous référant au modèle COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987). Pour cela, nous comparons l'effet de trois catégories de groupes de consonnes (occlusive-liquide correspondant à des groupes tautosyllabiques / fréquents, occlusive-fricative et occlusive-occlusive constituant des groupes hétérosyllabiques / rares) en imposant des contraintes lexicales au matériel choisi. Ainsi, aucune des séquences mot + C utilisées dans l'expérience ne peut correspondre au début d'un mot dans la langue (en tout cas pas dans la base de données Brulex, Content et al., 1990). Si l'effet de la structure phonologique des séquences de phonèmes est répliqué, il sera alors possible d'affirmer que les effets mis en évidence (McQueen, 1998; Vroomen & de Gelder, 1999) ne constituent pas des artefacts de phénomènes de sélection lexicale sérielle.

1.1. Méthode

1.1.1. Sujets

Trente et un étudiants en second cycle de psychologie à l'Université Paris 5 - René Descartes ont participé à cette expérience soit volontairement, soit en échange de crédits d'enseignement. Ils étaient tous de langue maternelle française et n'avaient jamais souffert de troubles auditifs.

1.1.2. Matériel

Les stimuli sont des bisyllabes de structure C(C)VCCVC sans signification. La moitié d'entre eux commence par un mot d'une syllabe correspondant à la séquence de phonèmes initiale C(C)VC (par exemple /brytrin/, 'brute' + /rin/). Trente-quatre mots monosyllabiques ont été sélectionnés, dans la base de données BRULEX (Content et al., 1990), afin de respecter un certain nombre de critères structuraux et lexicaux. A l'exception de l'un d'entre eux, aucun ne contient une voyelle qui permettrait d'influencer le rattachement syllabique de la consonne suivante (Goslin, Content, & Frauenfelder, 1999). Nous avons donc évité d'utiliser des mots contenant les phonèmes /o/ ou /e/ qui sont ouverts dans une syllabe lourde (comme /o/ dans [rɔk], 'roc' et /e/ dans [sɛr], 'serres') mais pas dans une syllabe légère (comme /o/ dans [ro # kaj], 'rocaille' et /e/ dans [se # re], 'serrer'). Le seul mot du matériel contenant une voyelle de ce type (/o/ dans 'faute') n'a finalement pas été pris en compte dans les analyses qui sont présentées ici[35].

Les mots monosyllabiques sélectionnés se terminent tous par une consonne occlusive. Les non-mots auxquels ils sont intégrés présentent une structure C(C)VCCVC pour laquelle la suite CC médiane peut correspondre à une suite occlusive-liquide (tautosyllabique / fréquente, par exemple /gl/), occlusive-fricative (hétérosyllabique / rare, par exemple /gʒ/) ou occlusive-occlusive (hétérosyllabique / rare, par exemple /gd/, cf. Tableau 7). La fréquence des suites de consonnes médianes n'a pas été contrôlée ; les suites tautosyllabiques sont en moyenne plus fréquentes que les suites hétérosyllabiques (occl.-liq. vs. occl.-fric., $t(39) = 10.763$, $p < .001$; occl.-liq. vs. occl.-occl., $t(39) = 13.352$, $p < .001$). Les séquences occlusive-fricative et occlusive-occlusive ne présentent quant à elles pas de différence significative de leurs probabilités d'occurrence ($t(66) = 1.723$, $p > .1$). Les suites occlusive-liquide donnent lieu à un non-alignement entre frontières lexicale et phonologique alors que les deux autres catégories de groupes produisent un alignement de ces deux types de frontières.

Tableau 7 : Conditions expérimentales de l'Expérience 1.

	occlusive-liquide	**occlusive-fricative**	**occlusive-occlusive**
Exemple	/brytrin/	/brytsin/	/brytkin/
Classification du groupe de consonnes	tautosyllabique	hétérosyllabique	
Statut de la consonne finale du mot	attaque syllabique	coda syllabique	
Probabilité d'occurrence moyenne des groupes médians	88 (6.5)	44 (21.9)	38 (22.5)

Aucune des séquences C(C)VC + C initiales ne constitue un début de mot possible dans la base de données BRULEX (Content et al., 1990). Par exemple, le mot 'bague' peut être intégré dans les séquences /baglym/ (occlusive-liquide,), /bagʒym/ (occlusive-fricative) et /bagbym/

[35] Son insertion dans les analyses de variance ne modifie cependant pas les résultats obtenus.

(occlusive-occlusive). La séquence /bag/ constitue le début d'un certain nombre de mots dans la langue (bagarre, baguette, bagatelle, …). Par contre, aucune des séquences /bag/ + C ne correspond à un début de mot dans les trois conditions expérimentales comparées. Ni /bagl/, ni /bagʒ/, ni /bagb/ ne constituent le début d'un mot dans la langue[36]. Afin de nous assurer que l'ensemble des stimuli sélectionnés correspondait à cette contrainte, nous avons utilisé un script AWK similaire à celui qui avait été mis en œuvre pour étudier la fréquence des groupes de consonnes en position initiale de mot. Le script lisait un fichier contenant la liste des suites C(C)VC + C et recherchait la présence éventuelle de chaque suite en début de mot. A la fin du traitement, un fichier était créé qui contenait la liste des séquences données en entrée et le nombre de fois où chaque séquence avait été rencontrée en début de mot dans la base de données. Pour chaque séquence, on vérifiait que le nombre était bien égal à 0. Dans le cas contraire, un phonème différent était choisi et le script relancé sur les suites modifiées jusqu'à ce que l'ensemble des suites de la liste corresponde à des séquences ne constituant le début d'aucun mot dans la base de données. Trente-quatre autres stimuli ont été choisis dans lesquels aucun mot de la langue n'est enchâssé (par exemple /ʒalkɑ̃b/ ou /ʃylit/).

Les stimuli sélectionnés ont été enregistrés sur DAT par un locuteur masculin. Celui-ci était installé dans une cabine insonorisée et prononçait les stimuli à raison d'un non-mot de deux syllabes toutes les trois secondes. Chaque stimulus était prononcé trois fois de suite avec une intonation la plus neutre possible. Les stimuli ont ensuite été digitalisés sur 16 bits (échantillonnage à 16 kHz) par l'intermédiaire de l'entrée audio d'une carte son et stockés sur le disque dur d'un ordinateur compatible PC. Le fichier correspondant a alors été édité afin d'extraire le signal correspondant à chaque stimulus et de le sauvegarder dans un fichier individuel. Un exemplaire de chaque item a été sélectionné parmi les trois répétitions. Les stimuli ont ensuite été édités afin de mesurer la durée entre le début du stimulus et celui de l'occlusion de la consonne finale du mot.

1.1.3. Procédure

Nous avons utilisé la tâche de *Word-spotting* (Cutler & Norris, 1988; McQueen, 1996) dans laquelle les participants doivent repérer des mots monosyllabiques prononcés dans un non-

[36] Par attesté dans la langue, nous entendons désormais attesté dans la base de données Brulex (Content, et al., 1990). Il est évident que cette base de données ne recense pas l'ensemble des mots de la langue française. Elle fournit cependant un outil inestimable pour conduire nos recherches. Il est certain que de nombreux mots de la langue n'apparaissent pas dans Brulex mais l'on peut raisonnablement considérer que l'ensemble des mots recensés ➥

mot de deux syllabes (par exemple 'vague' dans /vagryʒ/). Les mots étaient toujours prononcés à l'initiale du non-mot. Les participants étaient assis sur une chaise, face à un bouton réponse connecté au port parallèle d'un ordinateur. Les participants étaient prévenus qu'ils allaient entendre des stimuli sans signification dont certains commenceraient par un mot de la langue (la consigne est retranscrite en Annexe 12, p.XXIII). Ils n'étaient pas informés des mots qui allaient apparaître dans le matériel. Lorsqu'ils identifiaient un mot en position initiale de non-mot, ils devaient appuyer le plus rapidement possible sur le bouton réponse. Dans le cas contraire, ils devaient attendre l'essai suivant qui commençait 3 secondes après la présentation du stimulus. Chaque fois qu'ils fournissaient une réponse manuelle indiquant la détection d'un mot en position initiale de non-mot, ils devaient dire à voix haute quel mot ils avaient identifié puis se préparer à l'essai suivant. L'expérimentateur vérifiait alors que le mot identifié était bien celui qui était attendu. Dans le cas contraire, le fichier de résultats était édité à la fin de la passation expérimentale pour y noter que la réponse était incorrecte.

Les stimuli étaient présentés à un niveau d'écoute confortable (environ 65dB). Les temps de réaction étaient mesurés depuis le début du stimulus et corrigés automatiquement par le logiciel de passation expérimentale en soustrayant de la latence de chaque temps de réaction la durée entre le début du stimulus et le début de l'occlusion correspondant à la consonne finale du mot (les résultats de ces mesures sont fournis en Annexe 13, p.XXIII). L'expérience commençait par une phase de familiarisation avec la tâche. Chaque sujet était soumis aux trois conditions expérimentales de contexte phonologique mais n'entendait chaque mot qu'une seule fois. Trois listes expérimentales ont donc été créées, chaque sujet se voyant attribuer aléatoirement l'une d'entre elles.

1.2. Résultats

Pour l'ensemble des expériences dans lesquelles nous avons eu recours à la tâche de *word-spotting*, deux critères d'exclusion des données expérimentales ont été choisis. Le premier porte sur la sélection des participants. Les données d'un sujet ne sont prises en considération dans l'analyse que s'il fournit une réponse manuelle *doublée* d'une identification correcte du mot attendu dans plus de la moitié des essais (sans préjuger du temps de réaction associé). Le second critère est appliqué après celui de sélection des participants et permet de sélectionner les items pertinents pour l'analyse. On considère qu'un item ne peut être intégré dans l'analyse que si plus

---

constitue le vocabulaire de base d'un locuteur adulte et que les mots supplémentaires sont pour la plupart ➥

de la moitié des sujets ont fourni pour cet item une réponse manuelle *doublée* d'une identification correcte du mot attendu. Pour chaque analyse de variance, deux analyses ont été effectuées, l'une sur les latences de détection (en ms) et l'autre sur les taux d'erreurs (en %). Dans l'ANOVA portant sur les latences de détection, les temps de réaction intégrés dans l'analyse ne devaient pas excéder la valeur du temps de réaction moyen ± deux écarts-type (SD pour *Standard Deviation*). Dans le cas contraire, ils n'étaient pas pris en compte dans l'analyse. Aucun des temps enregistrés ne passait en fait en-dessous de cette limite. Le critère consiste donc, dans l'ensemble des analyses qui ont été effectuées dans cette thèse, à supprimer les temps de réaction qui dépassent la valeur de la moyenne *plus* deux écarts-type. Dans les analyses qui portent sur les erreurs, on a considéré comme erreur toute absence de réponse mais aussi (1) toute réponse ayant été suivie d'une prononciation du mot ne correspondant pas à celui qui était attendu et (2) tout temps de réaction supérieur à la limite déterminée dans l'analyse des temps de réaction. Nous présentons pour chaque ANOVA les résultats obtenus dans l'analyse par *Sujet* (F_1) et dans l'analyse par *Item* (F_2).

1.2.1. Analyse globale

Les données de l'un des participants n'ont pas pu être intégrées dans l'analyse en raison d'un taux d'erreurs ou d'omissions supérieur à 50%. Deux items ('blague' et 'type') ont été supprimés de l'analyse en raison d'un taux trop important d'erreurs d'identification ou d'absence de détection de la part des sujets. Un autre item a été retiré de l'analyse en raison de la présence d'une voyelle qui, selon les régularités phonologiques du français, est apte à fournir des indices sur la structure syllabique de la séquence (/o/ dans 'faute'). Les temps de réaction supérieurs à 1430 ms n'ont pas été introduits dans l'analyse. Au total, 15 % des données n'ont pas été intégrées dans l'analyse des temps de réaction.

1.2.1.1. Temps de réaction

La variable *Liste* n'interagit pas avec la variable *Contexte*. Elle n'est donc pas introduite dans les analyses de variance suivantes. Celles-ci sont conduites avec le *Contexte* (C_3) comme variable intra-sujet (S * C_3) et intra-item (I * C_3). La latence de détection moyenne est de 702 ms (SD = 181 ms). On observe des temps de réaction en moyenne plus longs pour les groupes occlusive-liquide (722 ms) que pour les deux autres catégories de groupes (cf. Figure 28). Les latences obtenues pour les groupes occlusive-fricative sont quant à elles plus élevées (703 ms) que celles de la condition occlusive-occlusive (680 ms). L'effet global du contexte phonologique

relativement rares.

est significatif dans l'analyse par sujet ($F_1(2,58) = 4.085$, $p < .05$) mais pas dans l'analyse par item ($F_2(2,60) = 1.452$, $p > .1$). Du fait du statut similaire des groupes occlusive-fricative et occlusive-occlusive en ce qui concerne leur syllabation, nous avons approfondi l'analyse en comparant les latences de détection observées dans ces deux conditions. Celles-ci ne diffèrent pas significativement ($F_1(1,29) = 3.486$, $p = .072$; $F_2 < 1$). Les temps de détection enregistrés dans la condition occlusive-liquide (tautosyllabique) sont quant à eux significativement plus élevés ($F_1(1,29) = 4.762$, $p < .05$) que pour les deux autres types de groupes de consonnes (hétérosyllabiques). Cette différence n'est cependant pas significative dans l'analyse par item ($F_2(1,30) = 2.205$, $p > .1$).

Cette expérience ne confirme qu'en partie les résultats obtenus avec la même tâche par McQueen (1998) dans l'analyse des erreurs ou les données de temps de réaction observées par Vroomen & De Gelder (1999) dans une tâche de détection de phonèmes. Seuls les effets étudiés dans l'analyse par sujet sont en effet significatifs. Aucun des effets de l'analyse par item n'atteint le seuil de significativité statistique.

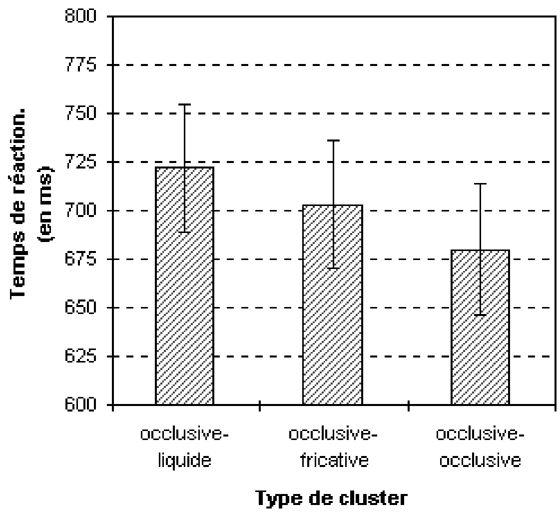

Figure 28 : Temps de réaction moyens (en ms) observés dans les trois conditions expérimentales de contexte phonologique de l'Expérience 1. Les barres verticales représentent l'erreur-standard[37] (SE).

1.2.1.2. Taux d'erreurs

Les mêmes analyses ont été conduites sur les taux d'erreurs (15 % de l'ensemble des données[38]). Le facteur *Liste* n'interagit pas avec le facteur *Contexte*. On observe un pourcentage d'erreurs légèrement plus élevé dans la condition occlusive-liquide (17.8 %) que dans les

[37] Erreur-standard (*Standard-Error*) : Ecart-type (*Standard Deviation*) divisé par le nombre d'observations (i.e. de sujets).

conditions occlusive-fricative (13.7 %) et occlusive-occlusive (14.4 %). L'effet du facteur *Contexte* n'est cependant significatif dans aucune des analyses ($F_1(2,58) = 1.380$, $p > .1$; $F_2 < 1$).

1.2.1.3. Discussion

Les données de temps de réaction répliquent donc en partie les effets observés par McQueen (1998) et par Vroomen & De Gelder (1999). Les locuteurs français ont plus de difficultés à détecter un mot en position initiale de non-mot lorsque la frontière entre le mot et la suite du non-mot est alignée avec une frontière pouvant s'assimiler à une frontière phonologique (syllabique ou phonotactique). Cet effet se maintient même si la suite mot + C ne constitue à aucun moment un début de mot dans la langue. Les effets observés précédemment ne sont donc pas uniquement interprétables à partir d'un modèle de sélection lexicale sérielle comme COHORT (Marslen-Wilson & Welsh, 1978 ; Marslen-Wilson, 1987). Il reste cependant que les effets observés dans cette expérience ne sont significatifs que dans les analyses conduites avec le facteur *Sujet* comme variable aléatoire. Toutes les analyses par item fournissent des résultats non-significatifs. Il semblerait donc que les latences de détection moyennées sur les items (les mots) lorsqu'on les regroupe par catégorie de contexte phonologique ne soient pas suffisamment stables pour générer des valeurs moyennes significativement différentes. Il est probable que ce phénomène puisse s'expliquer par des différences liées à des caractéristiques des mots choisis dans le matériel (par exemple leur fréquence d'usage). En effet, la tâche de *word-spotting* est particulièrement difficile à accomplir. Cette difficulté se manifeste aussi bien dans les jugements des participants que dans les taux d'erreurs et les latences de détection moyenne observés dans toute expérience faisant appel à ce paradigme (Cutler & Norris, 1988; McQueen, 1996). Du fait de cette difficulté, on peut imaginer que la présence de mots particulièrement difficiles à isoler sur la base de leurs caractéristiques intrinsèques pourrait donner lieu à des effets moins nets du contexte phonologique ; des mots relativement faciles à isoler donnant au contraire lieu à des effets plus clairs du contexte. La présence de mots pour lesquels le contexte n'a qu'un effet limité pourrait alors empêcher le test statistique de rejeter l'hypothèse nulle pour l'autre ensemble de mots. Nous avons donc choisi de scinder le matériel expérimental à partir de critères propres aux mots de notre échantillon. Si c'est effectivement ce paramètre qui conduit à une absence de significativité dans l'analyse par item, ceci devrait nous permettre de stabiliser la variabilité propre aux items et de faire émerger des comparaisons significatives dans cette partie de l'analyse.

---

[38] Les temps de réaction supérieurs à la moyenne + deux écart-types sont considérés comme des erreurs.

1.2.2. Analyses complémentaires

1.2.2.1. Contexte phonologique et fréquence d'usage

Nous avons en premier lieu cherché à évaluer l'influence possible de la fréquence d'usage des mots de l'échantillon dans l'émergence de l'effet du contexte phonologique. Il serait tout à fait possible que ce paramètre puisse influencer non seulement les temps de réaction observés en *word-spotting* mais aussi l'effet des divers facteurs introduits dans l'analyse. C'est par ailleurs un paramètre immédiatement disponible dans la base de données BRULEX (Content et al., 1990). L'objet n'est pas ici de tester l'effet de la fréquence des mots sur les temps de réaction mais plutôt l'influence que peut avoir eu la *variabilité attachée à ce paramètre* sur la capacité du test statistique à faire émerger des effets de contexte phonologique significatifs. Notre but est donc de limiter la variabilité des latences de détection liée à la fréquence des mots afin de faciliter l'émergence d'effets significatifs dans les analyses par item. Nous avons estimé la médiane[39] de la distribution des fréquences d'usage[40] des mots de notre échantillon afin de scinder celui-ci en deux groupes de taille approximativement égale sans nous préoccuper de la signification de la valeur de fréquence d'usage en termes de fréquence absolue d'utilisation dans la langue. La médiane est à 19. Les mêmes analyses ont alors été conduites sur chacun des échantillons de mots, l'un contenant les mots les plus rares de l'échantillon (FU < 19, N = 15, moyenne = 11), l'autre les plus fréquents (FU ≥ 19, N = 16, moyenne = 132). Puisqu'aucun effet, même marginal, n'émergeait dans l'analyse des erreurs, nous n'avons répliqué que les analyses de temps de réaction.

1.2.2.1.1. Résultats

Le facteur *Liste* n'interagit pas avec le facteur *Contexte* ; les analyses présentées ne le prennent donc pas en compte. Les latences de détection obtenues pour les 15 mots les plus rares de l'échantillon sont en moyenne de 713 ms (SD = 201 ms) et les moyennes observées ne correspondent pas à ce qui était obtenu dans l'analyse globale. Les groupes occlusive-fricative semblent donner lieu à des temps de réaction plus longs (741 ms) que les occlusive-liquide (708 ms) et les occlusive-occlusive (691 ms). On n'observe cependant aucun effet du contexte phonologique pour ce groupe de mots ($F_1(2,58) = 1.983$, $p > .1$; $F_2 < 1$). L'analyse conduite sur

---

[39] Médiane : cet indice statistique est la valeur qui sépare un échantillon en deux groupes de taille égale. La moitié des mots de l'échantillon a donc une fréquence d'usage supérieure à la médiane. Cet indice est ici plus intéressant que la moyenne puisque celle-ci peut séparer deux groupes de taille tout à fait inégale en raison du poids des valeurs extrèmes dans le calcul de la moyenne.

[40] Notée FU (Fréquence d'Usage sur 1 million). Les données de fréquence présentes dans BRULEX (Content et al., 1990) doivent être divisées par 100 pour aboutir à une fréquence d'usage sur 1 million.

les 16 mots les plus fréquents de l'échantillon (RT = 673 ms, SD = 198 ms) fournit des données sensiblement plus proches de ce qui était observé dans l'analyse globale. Les groupes occlusive-liquide donnent lieu à des latences de détection plus longues (720 ms) que les deux autres catégories de groupes (respectivement 662 ms et 637 ms pour les occlusive-fricative et les occlusive-occlusive). L'effet global du contexte n'est à nouveau significatif que dans l'analyse par sujet ($F_1(2,58) = 3.462$, $p < .05$) et pas dans l'analyse par item ($F_2(2,30) = 1.599$, $p > .1$). Nous avons à nouveau regroupé les groupes occlusive-fricative et occlusive-occlusive (qui ne diffèrent pas entre eux : $F_1 < 1$; $F_2 < 1$) et les avons comparés aux groupes occlusive-liquide. On observe de même que dans l'analyse globale un effet significatif par sujet ($F_1(1,29) = 7.838$, $p < .01$) mais celui-ci n'est que marginal dans l'analyse par item ($F_2(1,15) = 4.301$, $p = .056°$).

1.2.2.1.2. Discussion

La restriction à une partie du matériel linguistique utilisée dans l'Expérience 1 semble avoir permis au test statistique de rejeter plus facilement l'hypothèse nulle dans l'analyse par item, ce qui conduit à observer un seuil statistique marginal alors qu'il était non-significatif dans l'analyse globale. Cet effet n'est cependant que marginal. Nous nous sommes alors posé la question de l'existence d'une autre variable lexicale qui aurait pu introduire une variabilité dans les temps de réaction observés et serait plus appropriée que la fréquence d'usage des mots. Nous pensons que le nombre de compétiteurs lexicaux pourrait constituer une information influençant les traitements effectués au cours de la tâche de *word-spotting*. Nous en sommes arrivé à choisir cette variable en raison du lien qu'elle peut avoir avec la fréquence d'usage. En effet, (Luce, 1990) montre que les mots fréquents ont en général un nombre de voisins lexicaux plus important que les mots rares et que les effets de fréquence peuvent se confondre avec les effets de voisinage lexical. Fréquence et nombre de compétiteurs sont donc liés, ce qui nous conduit à envisager que si la conduite de l'analyse sur les mots fréquents contribue à observer un seuil de probabilité plus bas (bien que toujours supérieur au seuil de significativité), il pourrait en être de même si l'on restreint l'analyse aux mots ayant beaucoup de compétiteurs. Par ailleurs, cette variable nous semblait mieux à même d'interagir réellement avec le contexte phonologique. En effet, Norris, McQueen, & Cutler (1995) citent un travail non-publié dans lequel seuls des effets de fréquence relativement importants auraient été observés dans une tâche de *word-spotting*. Dans ce même travail, des différences de fréquence limitées ne permettent pas de faire émerger des effets significatifs. Or la scission que nous avons effectuée dans notre matériel ne nous permet pas de dissocier groupes très rares et groupes très fréquents. En réalité, la plupart des mots que nous avons utilisés appartiennent à la classe des mots de fréquence d'usage moyenne (la moitié des mots utilisés se situent entre 11 et 52 sur 1 million sur l'échelle de fréquence d'usage de BRULEX). La comparaison de mots présentant des différences relativement peu importantes en termes de fréquence ne devrait donc pas, dans une tâche de *word-spotting*, donner lieu à des différences importantes en termes de temps de réaction. Dans cette expérience nous avons pu observer, au cours de l'analyse ayant consisté à restreindre l'ANOVA aux mots fréquents d'une part et rares de l'autre, des temps de réaction moyens de 673 et 713 ms respectivement. Bien que cette différence soit significative ($t(29) = 2.099$, $p < .05$), il n'est pas certain qu'elle soit réellement liée à une différence de fréquence des items comparés. D'autres paramètres peuvent très bien être corrélés et déterminer une diminution des temps de réaction pour les mots fréquents sans que la différence soit effectivement liée à cette variable. On peut par exemple envisager que Si la fréquence d'usage n'est pas réellement en mesure d'influencer les traitements effectués au cours de la tâche, il est probable que la dichotomie réalisée ne soit pas à même de contribuer à faire émerger des effets significatifs pour une partie des items. Il nous semble plus probable que le nombre de compétiteurs d'un mot puisse influencer les processus qui sont à l'œuvre dans une tâche de *word-spotting*. En effet, la fréquence d'un mot n'a pas nécessairement de représentation dans le lexique. Son intervention pourrait d'ailleurs être envisagée en termes décisionnels ou intégratifs, c'est à dire qu'elle influencerait l'étape de prise de décision mais pas celle d'accès au lexique proprement dite. Le nombre de compétiteurs d'un mot constitue par contre une variable essentielle des processus de reconnaissance de la parole. Ainsi, dans le cadre d'un modèle comme COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987), ce paramètre est partie intégrante du traitement effectué sur le signal de parole. Le processus de reconnaissance des mots repose en effet essentiellement sur la sélection d'un mot parmi l'ensemble des candidats lexicaux activés dans la cohorte. Des modèles comme TRACE (McClelland & Elman, 1986) ou SHORTLIST (Norris, 1994) reprennent ce principe de sélection parmi un ensemble de candidats lexicaux en implémentant des processus de compétitions lexicales. Or ce type de processus fait implicitement intervenir le nombre de candidats lexicaux dans le déroulement des processus de traitement. Ce paramètre est donc en mesure d'influencer la durée nécessaire pour isoler un mot parmi l'ensemble des mots possibles et est devenu une caractéristique essentielle de tout modèle de l'accès au lexique. Il semble donc possible de choisir un critère alternatif à la fréquence pour tenter de réduire la variance liée aux items et de faire émerger un effet d'alignement significatif dans une analyse par item.

1.2.2.2. Contexte phonologique et nombre de compétiteurs

Afin d'affiner notre analyse du rôle des contraintes phonologiques dans les processus de segmentation de la parole en mots, nous avons donc choisi de conduire une analyse similaire à la précédente en choisissant un critère plus adéquat pour restreindre l'étude à une partie du matériel : le nombre de compétiteurs lexicaux des mots utilisés dans l'expérience.

1.2.2.2.1. Méthode

L'estimation du nombre de compétiteurs des items expérimentaux a été réalisée avec le même script AWK que celui qui nous avait permis de vérifier qu'il n'existait pas de séquence 'mot' + C constituant le début d'un mot lors de la constitution du matériel de l'Expérience 1. Le script prenait en entrée la liste des mots utilisés dans cette expérience et recherchait les occurrences de la séquence de phonèmes correspondante au début des mots de la base de données BRULEX (Content et al., 1990). Le fichier de sortie contenait la liste des mots et le nombre de mots commençant par cette séquence dans la base de données. Cette opération nous a permis de calculer la médiane de cet indice et de séparer les items en 2 groupes. Les analyses de variance conduites sur l'intégralité du matériel de l'Expérience 1 ont été reconduites sur chacun de ces groupes de mots en se restreignant à l'analyse des temps de réaction.

La médiane de la distribution des items en terme de nombre de compétiteurs est à 6. Or trois des mots de l'échantillon ont exactement 6 compétiteurs dans la base de données BRULEX (Content et al., 1990). Cette valeur pose un problème. Soit l'on décide de dissocier les résultats obtenus pour les mots ayant plus de 6 compétiteurs et les autres (On compare alors les résultats obtenus dans des groupes de 16 et 15 items respectivement pour les mots ayant peu et beaucoup de compétiteurs), soit l'on choisit d'intégrer dans le groupe des mots ayant beaucoup de compétiteurs ceux qui ont plus de 5 compétiteurs (donc 6 ou plus de 6) et de restreindre l'autre groupe aux items restants, ce qui conduit à comparer des échantillons de 13 et 18 items respectivement pour les groupes de mots ayant peu et beaucoup de compétiteurs. Les mots que nous avons qualifiés de fréquents ont en moyenne un nombre de compétiteurs plus élevé (10.06) que les mots rares (6.44). Cette différence n'est pas significative ($t(28) = 1.500$, $p > .1$) mais elle nous incite à privilégier l'une des deux possibilités qui s'offrent à nous pour la dichotomie que nous souhaitons effectuer. C'est en effet pour les mots fréquents que l'on a observé la possibilité d'une différence statistiquement significative entre les conditions de contexte phonologique dans la section précédente. Du fait de la quantité relativement restreinte de mots dans notre matériel, nous avons choisi de sélectionner une méthode de dichotomie qui nous permettrait de conserver au test statistique une puissance raisonnable lui permettant de faire émerger des effets significatifs. Cette analyse restreinte en fonction du nombre de compétiteurs des mots de l'échantillon a donc été conduite d'une part sur les mots qui ont 6 compétiteurs ou plus (18 items), et d'autre part sur ceux qui en ont moins de 6 (les 13 autres items). En raison du nombre de compétiteurs sensiblement plus élevé pour les mots fréquents que pour les mots rares, il est probable qu'un éventuel effet d'alignement devrait se manifester pour les mots ayant beaucoup de compétiteurs.

1.2.2.2.2. Résultats

Aucune des analyses effectuées ne fait ressortir un effet d'interaction entre les variables *Contexte* et *Liste* ; cette dernière n'est donc pas intégrée dans les analyses de variance présentées. De même que pour la restriction aux mots les plus rares, l'analyse des 13 mots pour lesquels le nombre de compétiteurs est le moins élevé ne permet pas de mettre en évidence des différences de temps de réaction entre les diverses conditions expérimentales. Les conditions occlusive-liquide, occlusive-fricative et occlusive-occlusive donnent lieu à des latences de détection qui sont respectivement de 690, 726 et 721 ms (tous les F sont inférieurs à 1, la latence moyenne de détection est de 712 ms, SD = 197).

L'analyse conduite sur les 18 mots de l'échantillon présentant la quantité la plus élevée de compétiteurs aboutit au contraire à des résultats très similaires à ceux que l'on a observé pour les mots les plus fréquents. Le temps de réaction moyen est de 698 ms (SD = 201 ms). Les groupes occlusive-liquide donnent lieu à des latences de détection plus longues (750 ms) que les groupes occlusive-fricative (691 ms) et occlusive-occlusive (655 ms). La Figure 29 fournit une représentation graphique de ces temps de réaction. L'effet global du contexte est significatif par sujet ($F_1(2,58) = 11.693$, $p < .01$) mais pas par item ($F_2(2,34) = 2.076$, $p > .1$). La comparaison des conditions occlusive-fricative et occlusive-occlusive montre que les temps de réaction

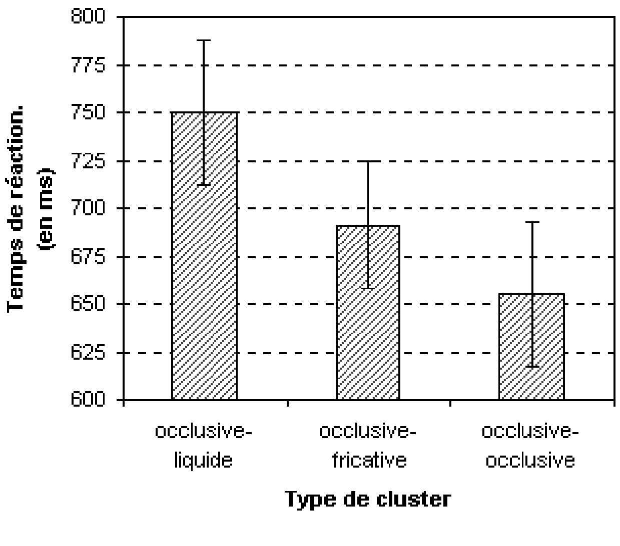

Figure 29 : Temps de réaction moyens (en ms) observés pour les 18 mots de l'échantillon ayant le nombre de compétiteurs le plus élevé. Les barres verticales représentent l'erreur-standard (SE).

associés ne diffèrent que marginalement dans l'analyse par sujet ($F_1(1,29) = 3.950$, $p = .056$) et pas dans l'analyse par item ($F_2 < 1$). Nous avons donc regroupé ces deux conditions et conduit une analyse permettant de comparer les groupes occlusive-liquide à ces deux autres catégories. Cette comparaison met en évidence des différences significatives aussi bien dans l'analyse par sujet ($F_1(1,29) = 19.614$, $p < .01$) que dans l'analyse par item ($F_2(1,17) = 4.643$, $p < .05$). Les mots sont détectés plus rapidement lorsque la consonne finale occlusive est suivie d'une liquide donnant lieu à un groupe de consonnes tautosyllabique / fréquent que lorsque cette même consonne est suivie d'une fricative ou d'une occlusive qui, quant à elles, génèrent un groupes de consonnes hétérosyllabique / rare.

1.3. Discussion

L'objet de cette expérience était de tester le rôle d'éventuelles connaissances sur les contraintes phonologiques de la langue avec un matériel permettant de neutraliser la contribution des processus de sélection sérielle dans l'émergence des effets d'alignement phonologique. Pour cela, nous avons utilisé des stimuli pour lesquels l'occurrence de la consonne contextuelle, celle qui suit le mot dans le matériel de *word-spotting*, permettait de mettre fin à toute compétition entre des candidats qui resteraient maintenus dans la cohorte lexicale initiale du modèle COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987). Pour l'ensemble des stimuli, la séquence 'mot' + C ne pouvait ainsi en aucun cas constituer le début d'un mot dans la base de données lexicale BRULEX (Content et al., 1990). Le choix de ce matériel rend donc impossible l'existence d'un déséquilibre du nombre de candidats lexicaux maintenus dans la cohorte à l'instant auquel est traitée la consonne contextuelle dans les conditions d'alignement et de non-alignement des frontières syllabique et lexicale.

Une première analyse a permis de répliquer l'effet du contexte phonologique, c'est à dire de l'alignement entre frontières syllabique et lexicale, malgré ce contrôle des processus de sélection lexicale sérielle. Cet effet n'a cependant pu être mis en évidence que dans l'analyse par sujet, celui-ci n'atteignant pas le seuil de significativité statistique dans l'analyse par item. La restriction de l'analyse de variance à une partie du matériel lexical nous a cependant permis de mettre en évidence un effet de la structure phonologique des groupes de consonnes sur les latences de détection enregistrées dans la tâche de *word-spotting*. Deux questions importantes se posent néanmoins à partir de ces résultats. Nous avons vu que l'effet du contexte phonologique n'est en réalité significatif que pour une partie des items expérimentaux : les mots ayant beaucoup de compétiteurs. Cette restriction de l'effet à une sélection de mots pose d'autant plus le problème de l'interprétation qu'il convient de donner de ces résultats. Nous avons vu, lors de la description du travail de McQueen (1998), que les effets contextuels observés dans la tâche de *word-spotting* pourraient très bien s'expliquer par des caractéristiques intrinsèques aux mots que doivent détecter les participants. Ce problème se pose lors de toute étude de phénomènes contextuels impliquant la présentation d'exemplaires physiques différents en fonction des conditions de contexte. La limitation de l'effet à une partie des mots du matériel accentue la possibilité que l'effet de contexte observé soit en réalité lié à certaines caractéristiques de la portion acoustique qui correspond aux mots utilisés. Il est donc essentiel de déterminer la contribution de caractéristiques intrinsèques aux mots dans l'émergence même de l'effet avant de réfléchir aux raisons pour lesquelles un éventuel effet des contraintes phonologiques pourrait se restreindre à une partie des mots de l'échantillon.

2. Expérience 2 : Contrôle en décision lexicale

L'objet de la seconde expérience est d'évaluer la contribution de paramètres acoustiques propres aux exemplaires des mots utilisés dans la tâche de *word-spotting* à l'émergence de l'effet d'alignement observé dans l'Expérience 1. A cet effet, nous avons mis en œuvre une tâche de décision lexicale sur les fragments acoustiques correspondant à la portion C(C)VC initiale des stimuli utilisés dans la précédente expérience. Si l'effet contextuel observé précédemment peut s'expliquer par des caractéristiques acoustiques propres aux exemplaires utilisés dans la première expérience et n'a rien a voir avec le statut de la consonne contextuelle, les effets observés dans l'Expérience 1 devraient également émerger dans une tâche de décision lexicale dans laquelle les conditions contextuelles n'interviennent pas en tant que tel. Ainsi, les mots extraits des stimuli de la condition occlusive-liquide devraient alors donner lieu à des latences de décision lexicale plus longues que les mots extraits des conditions occlusive-fricative ou occlusive-occlusive. Pour cela, nous comparons les réponses fournies pour chaque portion C(C)VC initiale en fonction du contexte dont elle est extraite. On observe par exemple les réponses induites par le mot 'brute' en fonction du stimulus dans lequel il était produit dans l'Expérience 1 (occlusive-liquide, occlusive-fricative, occlusive-occlusive) alors que seule la portion lexicale est disponible. Si la prononciation de certaines consonnes contextuelles a influencé la prononciation du mot initial au point de faire émerger les effets que nous avons observés dans la tâche de *word-spotting*, ces effets devraient se répliquer malgré l'absence de contexte. Au contraire, si l'effet n'est pas reproduit, il sera possible d'affirmer que les différences observées dans l'expérience précédente n'étaient pas liées à des différences acoustiques des stimuli qui seraient déterminées par les caractéristiques de la consonne contextuelle. Si l'effet observé dans l'expérience précédente est effectivement lié à la mise en œuvre de processus de segmentation lexicale reposant sur le statut de la consonne contextuelle, lequel influencerait l'alignement entre frontières phonologique-fréquentielle et lexicale, nous nous attendons à ne pas observer d'effet de l'origine des stimuli dans cette expérience.

2.1. Méthode

2.1.1. Sujets

Trente étudiants en second cycle de psychologie à l'Université Paris 5 - René Descartes ont participé à cette expérience, soit volontairement, soit en échange de crédits d'enseignement. Ceux-ci n'avaient pas pris part à la première expérience. Ils étaient tous de langue maternelle française et n'avaient jamais souffert de troubles auditifs.

2.1.2. Matériel

Les stimuli de l'Expérience 1 ont été édités avec un logiciel de traitement du signal. La fin acoustique de la portion correspondant à la séquence C(C)VC initiale a été estimée à partir du spectrogramme. Cette décision a ensuite été confirmée auditivement afin de s'assurer que la consonne contextuelle n'était effectivement pas audible jusqu'à ce point. La section comprise entre le début du stimulus et la fin acoustique de la portion C(C)VC a finalement été extraite et stockée sous le même format que les stimuli de l'Expérience 1 (digitalisation sur 16 bits, échantillonnage à 16 kHz). Cette opération a été effectuée avec les stimuli composés d'un mot en position initiale aussi bien qu'avec les stimuli de remplissage. Pour les stimuli-test, nous disposions alors de 3 exemplaires de chaque mot en fonction du contexte duquel il avait été extrait. Par exemple, la portion correspondant à la séquence /bag/ ('bague') a été extraite des stimuli /baglym/ (occlusive-liquide), /bagʒym/ (occlusive-fricative) et /bagbym/ (occlusive-occlusive). Les trois stimuli obtenus contiennent uniquement la séquence de phonèmes /bag/, mais celle-ci provient dans chacun d'entre eux de 3 exemplaires acoustiques différents. La même opération a été réalisée avec les stimuli de remplissage.

2.1.3. Procédure

La procédure de passation expérimentale est identique à celle de l'Expérience 1. Seule la tâche diffère. Les participants devaient effectuer une tâche de décision lexicale sur des stimuli mono-syllabiques. Les stimuli obtenus à partir de la portion initiale des items de remplissage de l'Expérience 1 correspondent à des non-mots. Ceux qui proviennent des stimuli-test forment des mots. Nous avons utilisé une procédure *go - no go* qui s'apparente à ce qui était demandé dans la tâche de word-spotting. Lorsque le sujet identifiait le stimulus comme étant un mot de la langue, il devait appuyer le plus rapidement possible sur le bouton réponse puis prononcer ce mot à voix haute. Dans le cas contraire, il n'appuyait pas et devait attendre l'essai suivant. L'intervalle inter-stimuli était le même que dans l'Expérience 1 (3 s). Les temps de réaction étaient mesurés depuis le début du stimulus et corrigés automatiquement par le logiciel de passation expérimentale en soustrayant de la latence de chaque temps de réaction la durée entre le début du stimulus et le début de l'occlusion correspondant à la consonne finale du mot. Cette durée est la même que celle qui était utilisée pour corriger les temps de réaction de l'Expérience 1. L'expérience commençait par une phase de familiarisation avec la tâche. Chaque sujet était soumis aux trois conditions d'origine du stimulus mais n'entendait chaque mot qu'une seule fois. Trois listes expérimentales ont été créées, chaque sujet se voyant attribuer aléatoirement l'une d'entre elles.

2.2. Résultats

Nous avons utilisé les mêmes critères de sélection des participants et des items que dans l'Expérience 1. Du fait de ces critères, les données de l'ensemble des participants ont été intégrées dans l'analyse. Cinq items ont cependant été supprimés de l'analyse en raison d'un taux trop important d'erreurs d'identification ou d'absence de détection de la part des sujets. Les temps de réaction supérieurs à 932 ms n'ont pas été introduits dans l'analyse. Les données non-intégrées dans l'analyse des latences de détection constituent 13 % de l'ensemble des observations.

2.2.1. Temps de réaction

Les temps de réaction et les taux d'erreur observés sont présentés dans le Tableau 8. La variable *Liste* n'interagit pas avec la variable *Origine*. Elle n'est donc pas introduite dans les analyses de variance présentées. Celles-ci sont conduites avec l'*Origine* (O_3) comme variable intra-sujet ($S * O_3$) et intra-item ($I * O_3$). La latence de détection moyenne est de 469 ms (SD = 88 ms). On observe des temps de réaction qui sont en moyenne plus *courts* pour les mots provenant de la condition occlusive-liquide (455 ms) que pour les deux autres catégories de groupes. Les latences obtenues pour les mots issus des conditions de contexte occlusive-fricative et occlusive-occlusive sont respectivement de 481 ms et 472 ms.

Le fait d'observer, pour les stimuli prononcés en condition occlusive-liquide, des temps de réaction en moyenne plus rapides que pour les stimuli prononcés dans les conditions occlusive-fricative et occlusive-occlusive est déjà un point important pour affirmer que l'allongement des latences de détection observé dans la tâche de *word-spotting* ne peut s'expliquer par des paramètres acoustiques (liés à des phénomènes articulatoires) qui ralentiraient l'identification des mots lorsque ceux-ci sont suivis d'une consonne liquide par rapport à une situation dans laquelle la consonne suivante est une fricative ou une occlusive. Il est cependant intéressant d'analyser plus en détails les mesures obtenues dans cette tâche. L'effet global de l'origine des stimuli est marginal dans l'analyse par sujet ($F_1(2,58) = 3.118$, $p = .052$) mais n'est pas significatif dans l'analyse par item ($F_2(2,56) = 1.722$, $p > .1$). Du fait du statut similaire des groupes occlusive-fricative et occlusive-occlusive pour ce qui a trait à leur syllabation, nous avons approfondi l'analyse en regroupant ces deux modalités en une seule. On observe à nouveau une absence de différence significative entre ces deux conditions (tous les F sont inférieurs à 1). Par contre, les temps de réaction observés pour les stimuli qui étaient présentés à l'origine en condition occlusive-liquide sont significativement plus courts que ceux des deux autres modalités ($F_1(1,29) = 4.708$, $p < .05$). Cette différence n'est que marginale dans l'analyse par item ($F_2(1,28) = 3.543$, $p = .070$). On observe par conséquent un effet non-négligeable de l'origine des stimuli. Il est probable que cette différence dans les temps nécessaires à fournir une décision lexicale sur le statut des mots en fonction du contexte duquel ils ont été extraits est liée à des paramètres acoustiques qui seraient déterminés par des contraintes articulatoires ayant influencé la prononciation de l'occlusive finale du mot.

Tableau 8 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 2.

	occlusive-liquide	**occlusive-fricative**	**occlusive-occlusive**
Temps de réaction	455 (75)	481 (99)	472 (89)
Taux d'erreur	8.1	17.6	13.3

Cet effet est cependant l'inverse de celui qui a été obtenu dans l'Expérience 1. L'analyse des temps de réaction mesurés dans la tâche de décision lexicale ne permet donc pas de rendre compte de l'effet d'alignement entre frontières syllabique et lexicale. Il est cependant courant d'observer des phénomènes de compensation entre latences de décision et taux d'erreurs. Dans ces situations, on peut par exemple observer une correspondance entre rétrécissement des temps de réaction et accroissement des taux d'erreur. Ce phénomène illustre généralement une attitude des sujets qui consiste à tenter de répondre plus rapidement dans certaines situations, ces réponses plus rapides induisant des erreurs plus nombreuses. Ce phénomène est donc plutôt lié à des stratégies de réponse qui n'illustrent pas réellement un effet au niveau des processus de traitement mais peuvent s'expliquer en termes stratégiques. On peut néanmoins envisager que, dans la tâche de décision lexicale, certaines caractéristiques acoustiques des stimuli produits en condition occlusive-liquide aient incité les participants à répondre plus rapidement (par exemple, moins d'indices acoustiques sur l'identité du phonème contextuel). Les réponses plus rapides aux mots extraits de la condition occlusive-liquide pourraient donc induire des taux d'erreurs plus importants dans cette condition. Il est par conséquent nécessaire d'analyser d'éventuels effets dans les taux d'erreurs afin de vérifier que les différences de temps de réaction observées dans la première expérience ne se répliquent pas ici avec une autre variable dépendante.

2.2.2. Taux d'erreurs

Les mêmes analyses ont donc été conduites sur les taux d'erreurs (13 % de l'ensemble des données). Le facteur *Liste* n'interagit pas avec le facteur *Origine*. On observe un pourcentage d'erreurs nettement plus bas dans la condition occlusive-liquide (8.1 %) que dans les conditions occlusive-fricative (17.6 %) et occlusive-occlusive (13.3 %). Cette observation met en évidence une similarité avec l'effet qui est observé dans les temps de réaction mesurés dans la même tâche. L'effet du facteur *Origine* est statistiquement significatif dans l'analyse par sujet ($F_1(2,58) = 5.711$, $p < .01$) mais pas dans l'analyse par item ($F_2(2,56) = 2.322$, $p > .1$). Les taux d'erreur observés dans les conditions occlusive-fricative et occlusive-occlusive ne diffèrent cependant pas significativement ($F_1(1,29) = 1.771$, $p > .1$; $F_2 < 1$). La comparaison des taux d'erreur observés dans ces deux conditions et de ceux de la condition occlusive-liquide montre que les erreurs sont significativement plus nombreuses lorsque les mots sont extraits d'un contexte hétérosyllabique que lorsqu'ils ont été prononcés dans un contexte tautosyllabique ($F_1(1,29) = 13.865$, $p < .01$). Cet effet n'est à nouveau que marginal dans l'analyse par item ($F_2(1,28) = 3.679$, $p = .065$). Les participants ont donc éprouvé moins de difficultés à fournir une décision sur le statut lexical des mots présentés lorsque ceux-ci étaient extraits des stimuli de la condition occlusive-liquide utilisés dans l'Expérience 1 que lorsque ceux-ci étaient extraits des stimuli des conditions occlusive-fricative et occlusive-occlusive. Cet effet s'exprime aussi bien dans les latences de détection que dans les taux d'erreur. Le sens de l'effet est cependant l'inverse de celui qui est observé dans l'Expérience 1 avec la tâche de word-spotting.

2.3. Discussion

Dans l'Expérience 1, l'effet d'alignement entre frontières syllabique et lexicale se manifeste malgré les contraintes lexicales que nous nous sommes imposées pour le choix du matériel expérimental. En effet, les stimuli que nous avons utilisés ne permettaient pas l'émergence d'un effet qui serait uniquement déterminé par un déséquilibre dans les quantités de candidats lexicaux maintenus dans la cohorte selon que le groupe formé par la consonne finale du mot et par la consonne qui lui succède est tautosyllabique / fréquent ou hétérosyllabique / rare. Ainsi, l'allongement des latences de détection observé lorsque frontières phonologique et lexicale sont discordantes ne peut s'expliquer intégralement à partir de processus qui seraient déterminés par un déséquilibre des phénomènes de sélection lexicale sérielle dans les conditions expérimentales comparées. Par ailleurs, cet effet n'est pas reproduit tel quel dans une expérience contrôle de décision lexicale dans laquelle les participants doivent donner une réponse concernant le statut lexical de la portion C(C)VC initiale extraite des stimuli utilisés dans la tâche de *word-spotting*. Au contraire, les mots extraits des stimuli de la condition occlusive-liquide présentent une tendance à être plus facilement identifiés que ceux extraits des deux autres conditions expérimentales.

Cette observation fait cependant émerger un certain nombre de questions concernant les données observées dans la tâche contrôle (8.1 %, 17.6 % et 13.3 % d'erreurs pour les groupes occl.-liq., occl.-fric. et occl.-occl. et des latences de détection de 455 ms, 481 ms, 472 ms) et la signification de celles qui sont obtenues dans l'Expérience 1 (pas d'effet sur les taux d'erreur ; un effet sur les temps de réaction : 750 ms, 691 ms, 655 ms respectivement pour les groupes de consonnes occl.-liq., occl.-fric. et occl.-occl.). On peut envisager que l'accroissement des taux d'erreur et des latences de décision lexicale observé dans la seconde expérience soit lié à une altération acoustique des stimuli qui se serait produite lors de l'extraction des portions initiales effectuée pour la constitution du matériel. On pourrait alors envisager que cette altération ait pu masquer un effet acoustique qui aurait le même sens que l'effet d'alignement mis en évidence dans la tâche de *word-spotting*. Il serait alors impossible d'affirmer que l'effet d'alignement de l'Expérience 1 est réel ; il pourrait tout simplement constituer un effet acoustique lui-même masqué par un autre effet se manifestant dans le sens inverse. Nous ne pensons cependant pas que l'effet de l'expérience contrôle puisse s'expliquer par une altération des portions acoustiques correspondant au mot dans les stimuli des conditions occlusive-fricative et occlusive-occlusive, celle-ci ayant eu lieu lorsque nous procédions à l'extraction de la séquence C(C)VC initiale. Un certain nombre de travaux nous incitent en effet à penser que ces deux catégories de groupes devraient au contraire être plus faciles à segmenter à partir d'une représentation acoustique du signal. Lors de la constitution des stimuli de l'Expérience 2, il nous a semblé plus difficile de séparer à l'aide du logiciel de traitement du signal la portion acoustique correspondant au mot lorsque celui-ci était suivi d'une liquide que lorsqu'il était suivi d'une fricative ou d'une occlusive. On trouve dans la littérature un certain nombre de travaux réalisés dans la mouvance du courant de la *phonologie articulatoire* (Browman & Goldstein, 1992) qui montrent que les groupes de consonnes tautosyllabiques sont réalisés par le biais de processus de contrôle articulatoire tout à fait différents de ceux qui sont à l'œuvre pour la production des groupes hétérosyllabiques. Browman & Goldstein (1988) ont étudié divers indices dérivés de données obtenues en cinéradiographie et ont comparé des groupes de consonnes identiques mais dont la structure syllabique variait. Ils ont par exemple observé des données articulatoires pour un groupe comme /sp/ selon que celui-ci était prononcé à l'intérieur d'une syllabe ou dans deux syllabes différentes. Ils ont concentré leur étude sur des séquences $VC_1C_2VC_3$ intégrées dans des suites plus longues et ont mesuré, à partir de l'enregistrement des mouvements des articulateurs,

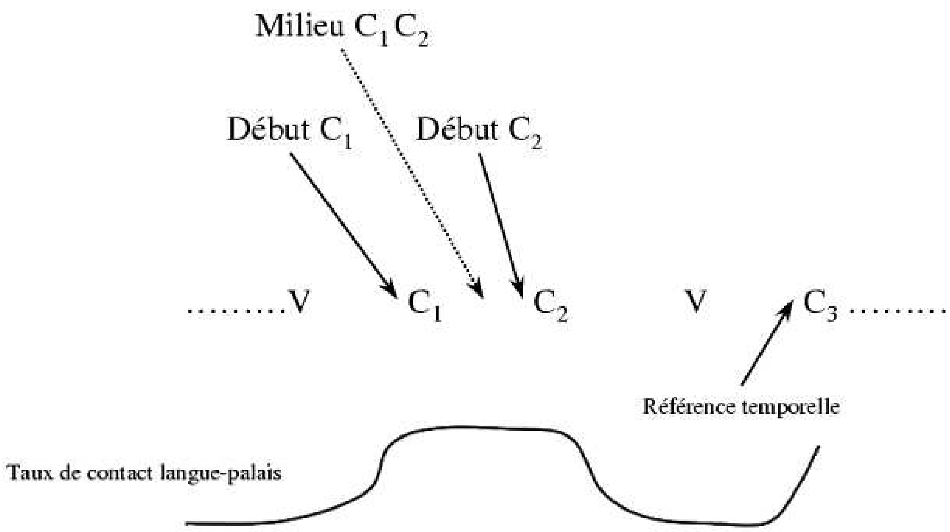

Figure 30 : Schéma représentant les résultats obtenus par Browman & Goldstein (1988) dans leur étude sur l'organisation articulatoire des séquences de consonnes en fonction de leur structure syllabique. La courbe du bas représente une mesure articulatoire des taux de contact entre langue et palais.

les durées qui séparent la consonne finale (C_3, la référence) des phonèmes qui constituent le groupe C_1C_2 (cf. Figure 30). A partir de ces données de durée, ils ont mesuré la variabilité des données en fonction du point de mesure (début de C_1, début de C_2, milieu de C_1C_2).

Ils observent que les durées séparant les groupes de consonnes hétérosyllabiques de la consonne suivante présentent une variance minimale lorsque l'intervalle temporel est mesuré entre le début de C_1 et la référence. Au contraire, l'intervalle observé pour les groupes tautosyllabiques présente une variance minimale lorsque la mesure est effectuée entre le milieu de la séquence C_1C_2 et la consonne de référence. Cette observation est interprétée par les auteurs comme un signe de la cohésion articulatoire élevée des groupes tautosyllabiques dans le cadre des processus de production de la parole. Cette observation incite à penser que la production d'un groupe de consonnes en attaque de syllabe passe par une représentation unitaire de la paire de consonnes alors que celle des groupes de consonnes hétérosyllabiques correspondrait à la juxtaposition de deux processus relativement indépendants pour chaque élément du groupe.

Les travaux de Byrd (1996) conduisent à des conclusions similaires à partir de l'analyse des recouvrements de gestes articulatoires en fonction du statut syllabique des séquences de consonnes. Les gestes correspondant à la production de chacun des phonèmes se recouvriraient beaucoup plus largement dans les groupes tautosyllabiques que dans le cas de la production de groupes hétérosyllabiques. Si ces phénomènes articulatoires ont aussi une manifestation acoustique spécifique, il serait certainement plus difficile de dissocier les composantes acoustiques correspondant aux segments de la séquence tautosyllabique que celles correspondant aux segments hétérosyllabiques. Ceci pourrait alors expliquer la difficulté que nous avons éprouvée à localiser la frontière entre les deux consonnes pour les stimuli de la condition occlusive-liquide. Cette difficulté a certainement influencé la possibilité d'extraire la portion correspondant au mot en évitant au maximum de supprimer des composantes acoustiques intervenant dans l'identification de celui-ci. Si c'est effectivement le cas, c'est à dire si les mots extraits de la condition occlusive-liquide ont été altérés par la procédure de découpage nécessaire à la mise en place de l'expérience contrôle et que l'effet observé dans la tâche de décision lexicale devait trouver sa source dans cette altération, nous aurions dû observer l'effet inverse : les stimuli extraits des séquences de la condition occlusive-liquide donnant lieu à des taux d'erreur plus importants et à des latences de détection plus longues. Il est donc peu probable que l'effet observé dans l'Expérience 1 puisse s'expliquer par des paramètres uniquement acoustiques qui auraient été masqués par une altération des stimuli extraits de la condition occlusive-liquide dans la tâche contrôle de décision lexicale.

Il est cependant possible de proposer une piste pour interpréter l'effet observé dans la tâche de décision lexicale. Il nous semble que les occlusives prononcées devant une liquide ont moins tendance à subir un phénomène qui pourrait s'apparenter à une réduction acoustique de la consonne. Ces phénomènes de réduction acoustique sont couramment observés en production spontanée. Ils consistent notamment, dans le cas des consonnes aussi bien que des voyelles, à modifier l'intensité et la composition spectrale de la portion acoustique subséquente à l'occlusion (Van Son & Pols, 1999). Ce phénomène se manifeste notamment lorsque les locuteurs sont dans une situation dans laquelle les processus de production de la parole sont moins contrôlés. Les caractéristiques d'une consonne réduite varient en fonction du type de consonne. Les fricatives et les occlusives réduites se caractériseraient par une diminution du rapport d'intensité consonne / voyelle. Cette réduction correspondrait également à une diminution du centre de gravité spectral pour les fricatives ainsi que pour les nasales (Van Son &

Pols, 1999). Cette réduction consonantique rend plus difficile l'identification de la consonne dans une tâche d'identification phonémique (Van Son & Pols, 1997). Lorsque nous avons procédé au découpage des stimuli afin d'en extraire les séquences C(C)VC initiales, il nous est apparu que l'occlusive finale des mots, qui est à l'initiale des groupes testés, était parfois difficilement visible sur la représentation spectrographique. Il est possible que la proportion de stimuli dans lesquels la consonne finale presque absente du signal soit plus importante parmi ceux des conditions occlusive-fricative et occlusive-occlusive que pour les items extraits de la condition occlusive-liquide. Nous n'avons pas conduit d'analyses acoustiques consistant à extraire l'enveloppe spectrale ou l'intensité moyenne de la section correspondant aux occlusives extraites des trois types d'environnements. En effet, quelle que soit l'interprétation à donner de l'effet observé en décision lexicale, son inversion par rapport à ce qui est observé dans l'Expérience 1 suffit à valider l'interprétation contextuelle proposée. Les travaux présentés par Browman & Goldstein (1988) et par Byrd (1996; Byrd & Tan, 1996) contribuent à penser qu'une altération liée à la méthode de découpage des stimuli aurait dû produire les effets inverses. S'il y a effectivement réduction consonantique plus marquée dans les stimuli occlusive-fricative et occlusive-occlusive, cette réduction est également présente dans les stimuli qui ont été utilisés dans l'Expérience 1 et n'est pas induite par la procédure d'extraction de la portion initiale des non-mots. L'éventuelle réduction consonantique aurait donc dû influencer également les réponses mesurées dans la tâche de *word-spotting*. Tout nous incite donc à penser que l'effet d'alignement observé dans la première expérience trouve bien sa source dans un effet de contexte qui est déterminé par le type de consonne qui suit la portion initiale du stimulus. Cet effet pourrait éventuellement -mais pas seulement- se décrire en termes d'alignement ou de non-alignement des frontières syllabique et lexicale. Nous sommes cependant confrontés à un effet qui se limite à une partie du matériel, ce qui nous incite à examiner plus en détails les relations qui peuvent exister entre des influences de type prélexical (structure phonologique, fréquence) et lexical (nombre de compétiteurs).

3. Discussion générale

Si l'effet d'alignement observé dans la première expérience n'est probablement pas lié à des caractéristiques acoustiques des stimuli, il reste que la restriction de l'effet à une partie du matériel doit être analysée plus en détails. Contrairement à ce que nous avions supposé pour justifier la séparation en deux groupes de mots, une différence dans la variabilité des résultats en fonction du type de mots ne semble pas justifier la restriction de l'effet à un sous-ensemble des items. En effet, on observe entre les mots qui ont beaucoup ou peu de compétiteurs des écart-types très similaires (respectivement 184 et 179 ms) alors que cette dichotomie contribue à faire émerger un effet significatif pour les mots qui ont beaucoup de compétiteurs. Il est donc probable que l'explication de cette restriction de l'effet ne soit pas statistique mais trouve sa source dans une relation particulière entre deux catégories d'informations qui sont supposées correspondre à deux niveaux de traitement des processus de segmentation de la parole en mots : un niveau prélexical à l'intérieur duquel interviendraient les connaissances sur les régularités phonologiques ou probabilistes de la langue et un niveau lexical dans lequel s'opèrent des phénomènes de compétition entre candidats lexicaux. Il serait alors délicat d'affirmer que l'effet observé découle de processus de traitement prélexicaux ayant pour objet de procéder à une segmentation du signal de parole préalable à la reconnaissance des mots. En effet, si la segmentation phonologique a pour fonction de faciliter la localisation des frontières de mots, elle doit intervenir avant tout recours à des processus de traitement lexicaux (compétition entre candidats notamment). Si un paramètre lexical (le nombre de compétiteurs) peut influencer le recours à un paramètre infralexical (la structure phonologique de la séquence), on est en mesure de proposer deux classes de modèles pour expliquer ce lien. La première possibilité est que la présentation d'une séquence de phonèmes correspondant à un mot qui a peu de compétiteurs dans le lexique va faciliter sa reconnaissance. Cette facilitation rendrait inutile le recours à des indices prélexicaux de segmentation. Une discussion plus approfondie de cette proposition et de ses implications sera présentée dans le Chapitre 7 en guise de réflexion sur les niveaux et les modes d'implémentation de processus reposant sur le recours aux contraintes phonologiques de la langue ; l'objet du Chapitre 6 est d'évaluer la validité d'un seconde proposition. Ce phénomène d'interaction entre variables lexicale et prélexicale pourrait en effet être conçu comme la mise en évidence du caractère lexical de ce que nous avons envisagé comme un effet prélexical. Il est en effet beaucoup plus probable d'observer un effet d'interaction entre des variables intervenant à un même niveau de traitement qu'à des niveaux différents.

Rappelons que le lien entre fréquence des groupes de consonnes et structure syllabique nous a conduits à proposer plusieurs interprétations de l'effet d'alignement. L'une d'entre elles permet d'adopter une conception purement lexicale de cet effet (cf. section 4.2.3 du Chapitre 4) selon laquelle un groupe tautosyllabique / fréquent déclencherait l'activation d'une quantité importante de candidats lexicaux. C'est cet accroissement du nombre de candidats lexicaux qui induirait un allongement des latences d'identification dans la tâche de *word-spotting*. Lorsque nous avons dichotomisé l'échantillon des mots utilisés dans l'Expérience 1, nous avons choisi une méthode de calcul du nombre de compétiteurs d'un mot qui relève d'un modèle de reconnaissance pour lequel les débuts de mots remplissent un rôle essentiel. En effet, nous avons procédé à ce calcul en recensant pour chaque mot les éléments du lexique commençant par cette séquence de phonèmes. La valeur que nous avons obtenue est donc un comptage du nombre de candidats qui sont maintenus dans cohorte initiale du modèle COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987) lorsque le phonème final du mot est en cours de traitement. Nous avons par conséquent choisi de manière arbitraire un mode de calcul qui accentue l'importance des débuts de mot. Lorsque nous avons envisagé que des groupes tautosyllabiques pourraient générer l'activation d'une quantité importante de candidats lexicaux en raison de leur fréquence d'occurrence élevée, nous avons fait référence à une conception différente de cette variable 'nombre de candidats lexicaux'. Cette dernière a alors été envisagée en rapport avec le modèle TRACE (McClelland & Elman, 1986), modèle pour lequel les phénomènes d'activation lexicale suivent une procédure d'alignement exhaustif sans que les débuts de mots se voient attribuer une importance particulière. Chaque phonème de la chaîne parlée pourrait alors générer l'activation d'un ensemble de candidats lexicaux quelle que soit la position du phonème dans les mots activés. Cette variable peut approximativement s'apparenter au 'nombre de voisins lexicaux' du *Neighborhood Activation Model* (NAM, Luce & Pisoni, 1998) et correspondrait au nombre d'unités lexicales qui subissent un accroissement de leur niveau d'activation dans le modèle TRACE (McClelland & Elman, 1986). Indépendamment du rôle essentiel de l'ordre des phonèmes, cet indice correspond au nombre de mots qui partagent 1 à N phonèmes avec le stimulus. Ainsi, dans le cadre d'un modèle à alignement initial, la séquence /vagryʒ/ donne lieu à l'activation de tous les mots commençant par /v/ ; ce qui correspond à 1040 mots dans la base de données BRULEX (Content et al., 1990). Le mot pertinent est alors progressivement sélectionné parmi les éléments de cette cohorte initiale. Cette sélection s'opère à partir d'un flux d'informations acoustico-phonétiques ascendant. Lorsque le phonème final /g/ est traité, il reste 15 éléments dans la cohorte de candidats. Dans le cadre d'un modèle à alignement exhaustif, le processus d'activation est très différent. De manière simplifiée, chaque phonème donne lieu à l'activation d'un ensemble indépendant de mots. Ces mots peuvent contenir le phonème en quelque position que ce soit. Ainsi, le /v/ de /vag/ activera tous les mots qui contiennent /v/ dans le lexique (3688 mots), puis tous ceux qui contiennent /a/ (14446 mots) et enfin les candidats qui intègrent le phonème /g/ (2940 mots). Les processus de compétition entre les divers candidats activés permettent d'aboutir à la reconnaissance du mot 'vague'[41]. Cette reconnaissance repose

[41] En réalité, la taille du lexique utilisé dans les simulations du modèle TRACE (McClelland & Elman, 1986 ; voir aussi Frauenfelder & Peeters, 1990) est beaucoup plus modeste, ceci afin de faciliter la procédure de simulation. ➥

-de même que dans COHORT- sur un flux ascendant d'informations acoustico-phonétiques mais aussi sur des phénomènes d'inhibition entre candidats lexicaux.

Même si cette procédure rend mieux compte de la reconnaissance des mots en parole continue en raison de l'existence de mots ayant un Point d'Unicité tardif (cf. Chapitre 2), le décours temporel du signal de parole incite cependant à ne pas abandonner totalement une composante séquentielle du traitement pour laquelle les activations lexicales seraient au moins partiellement alignées sur les débuts de mots. (cf. Mattys, 1997 pour une revue). Ces deux catégories d'alignement des activations pourraient donc constituer deux facettes du système de reconnaissance des mots en modalité auditive. On pourrait ainsi envisager un modèle dans lequel les procédures d'activation lexicale suivraient un alignement exhaustif mais où les candidats activés à partir d'un phonème présentant une forte probabilité de constituer un début de mot subiraient un accroissement de leur activation. Ceci accentuerait le poids du traitement séquentiel effectué sur le signal de parole tout en laissant la possibilité de faire intervenir des processus de reconnaissance fondés sur un alignement exhaustif afin de maintenir la possibilité de retarder la reconnaissance d'un mot s'il a par exemple un Point d'Unicité tardif.

On peut alors imaginer que les processus influencés par le nombre de compétiteurs du mot à partir de son début vont être sensibles à un accroissement du nombre de candidats activés en d'autres endroits du signal. Si l'on peut concevoir un modèle de la reconnaissance des mots dans lequel ces deux variables seraient relativement orthogonales, il devient possible d'envisager que la restriction de l'effet du type de groupe consonantique observée dans l'Expérience 1 trouve sa source dans une interaction entre deux variables lexicales plutôt qu'entre une variable lexicale et une variable infralexicale. L'effet d'alignement observé dans l'Expérience 1, de même que ceux mis en évidence par McQueen (1998) et par Vroomen & De Gelder (1999), pourrait alors se réduire à un effet du nombre de candidats lexicaux activés par le groupe de phonèmes constitué du phonème final du mot et du phonème contextuel. Cette interprétation remettrait en cause un modèle selon lequel les processus de segmentation lexicale peuvent avoir recours à des connaissances sur les régularités phonologiques de la langue.

Mais on peut considérer qu'un système de reconnaissance des mots reposant sur ces principes activerait approximativement cette quantité de mots.

Résumé

Les expériences présentées dans ce chapitre avaient pour objet de confronter l'effet d'alignement entre frontières phonologique et lexicale observé par McQueen (1998) et par Vroomen & De Gelder (1999) à une interprétation lexicale selon laquelle les deux types de conditions d'alignement correspondraient en réalité à un déséquilibre du nombre de compétiteurs maintenus dans la cohorte lorsque le phonème contextuel est en cours de traitement. L'Expérience 1 a été conduite avec un paradigme de *word-spotting* dans lequel l'occurrence de la consonne contextuelle donnait lieu à une rupture nette dans le processus de dé-activation des candidats de la cohorte. La séquence 'mot' + consonne contextuelle ne correspondait en effet au début d'aucun mot dans la base de données BRULEX (Content et al., 1990). L'effet d'alignement est répliqué pour une partie seulement des mots de l'échantillon. Une tâche contrôle de décision lexicale permet d'affirmer que l'effet observé n'est pas déterminé par des propriétés acoustiques des stimuli utilisés dans la tâche de *word-spotting*. La restriction de l'effet aux mots ayant beaucoup de compétiteurs incite cependant à envisager que cet effet soit lié à une variable lexicale plutôt qu'infralexicale.

Chapitre 6

Contraintes phonologiques, régularités probabilistes, compétitions lexicales

Ce chapitre a fait l'objet de deux communications orales dont une dans un congrès international avec comité de lecture assorti d'une publication dans des actes :

Eurospeech'99, ESCA 7th European Conference on Speech Communication and Technology, 5-10 Septembre 1999, Budapest, Hongrie.

Journée 'Langage et Lexique', 6 Novembre 1999, Institut des Sciences Cognitives, Lyon, France.

CONTRAINTES PHONOLOGIQUES, REGULARITES PROBABILISTES, COMPETITIONS LEXICALES

Les données obtenues dans l'Expérience 1 permettent de rejeter une interprétation selon laquelle l'effet d'alignement entre frontières phonologique et lexicale pourrait s'expliquer par un déséquilibre des phénomènes de sélection lexicale lors du traitement de la consonne contextuelle. L'effet observé se restreint cependant à une partie des mots de l'échantillon. Seuls les mots qui ont beaucoup de compétiteurs permettent en effet de faire émerger cet effet. Pour ce sous-ensemble de mots, les latences de détection sont plus élevées lorsque le mot à détecter est suivi d'une consonne liquide (laquelle donne lieu à un groupe de consonnes tautosyllabique / fréquent) que lorsque ce même mot est suivi d'une fricative ou d'une occlusive (le groupe de consonnes constitué étant alors hétérosyllabique / rare). Il semble donc possible de reproduire un effet d'alignement indépendamment d'un déséquilibre éventuel des activations lexicales séquentielles ; la restriction de cet effet aux mots qui ont beaucoup de compétiteurs incite à réfléchir sur la nature des processus qui conduisent à faire émerger un effet d'alignement. S'il est envisageable que cet effet soit de nature prélexicale, il est également possible que les processus qui déterminent son émergence soient localisés dans les étapes lexicales du traitement. Cette restriction de l'effet à une catégorie de mots soulève en effet la question du rôle de processus de compétition lexicale qui relèveraient de la fréquence des groupes de consonne médians. La fréquence des groupes de consonnes détermine directement le nombre de mots qui entreront en compétition avec le stimulus à identifier. L'occurrence d'une séquence occlusive-liquide (groupe tautosyllabique / fréquent) va activer une quantité plus importante de mots que celle d'une suite occlusive-fricative ou occlusive-occlusive qui sont hétérosyllabiques et rares. La relation entre structure syllabique et fréquence pourrait donc expliquer la restriction de l'effet d'alignement à certains mots. Cet effet d'alignement ne serait que la conséquence d'un processus de compétitions lexicales déterminé par la fréquence des groupes de consonnes comparés. Deux processus interagiraient donc dans la tâche : l'un d'entre eux impliquerait des processus de compétition entre candidats activés selon une méthode d'alignement exhaustif (McClelland & Elman, 1986) et s'assimilerait dans les tâches mises en œuvre à un effet d'alignement ; l'autre correspondrait à un mécanisme de sélection séquentielle des candidats lexicaux similaire à celui qui est proposé dans COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987). L'objet du Chapitre 6 est donc d'approfondir l'étude du rôle des contraintes phonologiques dans les processus de segmentation lexicale en contrôlant plus précisément le lien entre syllabation et fréquence.

Rappelons néanmoins qu'un effet de la fréquence des groupes de consonnes pourrait tout aussi bien se manifester à un niveau prélexical du traitement. Le système de segmentation lexicale utiliserait ainsi des connaissances sur les probabilités d'occurrence des suites de phonèmes pour localiser les frontières de mots (Brent & Cartwright, 1996; Saffran, Newport et al., 1996). Un effet de la fréquence des groupes de consonnes peut donc désigner deux classes de processus intervenant l'une au cours des étapes prélexicales (segmentation probabiliste) et l'autre au cours des étapes lexicales (compétitions entre candidats lexicaux). Il est ainsi possible qu'un effet de fréquence émerge sans que l'on puisse nécessairement conclure à un processus reposant sur des procédures lexicales. Notre objectif essentiel est cependant de déterminer dans quelle mesure on peut effectivement parler de segmentation fondée sur les contraintes phonologiques de la langue à partir des données observées en détection de phonèmes (Vroomen & de Gelder, 1999) ou en word-spotting (McQueen, 1998, cf. aussi l'Expérience 1 du Chapitre 5]. Les expériences présentées dans ce chapitre ont été construites afin de manipuler indépendamment fréquence des groupes et structure phonologique. Les données distributionnelles présentées dans le cadre du Chapitre 4 nous permettent de procéder à ce contrôle. Malgré une tendance moyenne des groupes tautosyllabiques à être utilisés plus fréquemment que les groupes hétérosyllabiques dans les mots de la langue, nous avons vu (Chapitre 4, Section 2.2.2) que les distributions de probabilité d'occurrence des groupes à initiale occlusive ou fricative se recouvrent. Il est donc possible de sélectionner des groupes de fréquence similaire qui ne présentent pas la même structure syllabique en position intervocalique. A l'inverse, les distributions de fréquence des groupes hétérosyllabiques présentent une étendue importante. Nous serons donc en mesure de comparer l'effet de la fréquence pour des groupes adoptant la même structure syllabique.

Nous étudierons donc d'une part le rôle de la structure phonologique à fréquence d'occurrence similaire et, d'autre part, le rôle de la fréquence en comparant des groupes de consonnes subissant une syllabation identique. Si les effets d'alignement observés sont effectivement liés à une segmentation fondée sur des contraintes phonologiques, nous devrions être en mesure de répliquer ces effets malgré un contrôle strict de la fréquence des groupes de consonnes. Si nous mettons en évidence un effet de la fréquence mais pas de la structure syllabique, il sera par conséquent difficile de départager une explication reposant sur un principe de segmentation prélexicale probabiliste d'une interprétation purement lexicale selon laquelle ce sont les processus de compétition entre candidats lexicaux qui détermineraient l'émergence de l'effet. Il sera cependant intéressant d'évaluer le rôle du nombre de compétiteurs des mots afin d'observer l'éventuelle présence d'une restriction de l'effet pour certaines variables. Il est évident que tous ces effets pourraient se manifester. L'émergence d'un effet de fréquence n'implique pas une absence d'effet 'phonologique' ; il est cependant intéressant de contraster les deux types de régularités afin de faire émerger au moins l'un des deux effets.

Nous utilisons à nouveau la tâche de *word-spotting* pour étudier le rôle des contraintes phonologiques dans les processus de segmentation lexicale en manipulant indépendamment la structure syllabique des groupes de consonnes et leur fréquence d'occurrence dans la langue. Par fréquence, nous entendons à nouveau ici la fréquence d'occurrence indépendamment de la position dans les mots (*probabilités brutes*). Notre objectif est ici de supprimer le rôle possible de la fréquence dans l'émergence de l'effet d'alignement. Parallèlement, nous étudions l'effet de la fréquence des groupes de consonnes en comparant des groupes présentant une structure phonologique similaire (Expérience 4).

1. Expérience 3 : Groupes de consonnes de même fréquence

Si les données de l'Expérience 1 permettent de réfuter une interprétation reposant sur un déséquilibre du nombre de candidats qui sont maintenus par les processus de sélection lexicale lorsque la consonne contextuelle donne lieu à un groupe tautosyllabique / fréquent ou hétérosyllabique / rare, nous ne sommes pas en mesure d'affirmer que l'effet d'alignement observé est effectivement lié à des processus qui reposeraient sur le recours à des connaissances concernant les régularités phonologiques de la langue. Du fait du lien mis en évidence dans le Chapitre 4 entre structure phonologique et fréquence, il est tout à fait possible que l'effet d'alignement entre frontières phonologique et lexicale constitue en réalité un effet d'alignement entre frontière probabiliste et lexicale tout autant qu'un effet lié à la sélection du candidat correct parmi un nombre plus élevé de candidats qui seraient activés selon une procédure l'alignement exhaustif. Il est donc nécessaire de contrôler précisément la fréquence des groupes de consonnes comparés afin d'isoler un effet purement phonologique. Pour cela, nous comparons à nouveau l'extraction de mots prononcés en position initiale de non-mot en fonction du type de groupe de consonnes médian. Ces groupes sont soit tautosyllabiques (occlusive-liquide) soit hétérosyllabiques (occlusive-fricative ou occlusive-occlusive). Contrairement au matériel de l'Expérience 1, leur fréquence d'occurrence dans la base de données Brulex (Content et al., 1990) est cependant maintenue constante dans les deux conditions de contexte phonologique. Ce contrôle est effectué à partir des données de probabilité d'occurrence qui ont été obtenues dans l'analyse distributionnelle du Chapitre 4. Du fait du contrôle strict de la fréquence des groupes de consonnes médians, un effet du contexte ne pourra émerger que si l'alignement entre frontières phonologique et lexicale est effectivement essentiel pour la segmentation du signal de parole en mots. Si au contraire les effets observés dans l'Expérience 1 ainsi que dans les travaux de McQueen (1998) et de Vroomen & De Gelder (1999) sont liés à un défaut d'alignement *entre frontières probabiliste et lexicale* ou à un déséquilibre du nombre de candidats lexicaux activés par l'occurrence du groupe de consonnes médian, ceux-ci devraient disparaître avec le matériel expérimental utilisé.

1.1. Méthode

1.1.1. Sujets

Quarante étudiants en second cycle de psychologie à l'Université Paris 5 - René Descartes ont participé à cette expérience, soit volontairement, soit en échange de crédits d'enseignement. Ceux-ci n'avaient pas pris part aux expériences précédentes. Ils étaient tous de langue maternelle française et n'avaient jamais souffert de troubles auditifs.

1.1.2. Matériel

Les stimuli sont des bisyllabes de structure C(C)VCCVC sans signification. La moitié d'entre eux (au nombre de 20) commence par un mot d'une syllabe correspondant à la séquence de phonèmes initiale C(C)VC (par exemple /baglyr/, 'bague' + /lyr/). L'objectif de cette expérience étant de contrôler la contribution de la fréquence des groupes de consonnes dans l'émergence de l'effet contextuel observé, nous avons été amenés à utiliser un nombre restreint de groupes consonantiques médians. Par conséquent, les mots utilisés dans le matériel ne sont pas les mêmes que ceux de l'Expérience 1. Nous avons par contre appliqué les mêmes critères de choix du matériel. Aucun des mots à détecter ne contient une voyelle qui permettrait d'influencer le rattachement syllabique de la consonne suivante (comme /o/ ou /e/). Les mots monosyllabiques sélectionnés se terminent tous par une consonne occlusive. Chaque mot est prononcé dans deux séquences sans signification correspondant chacune à deux conditions de contexte phonémique. Ces non-mots présentent une structure C(C)VCCVC dans laquelle la suite CC médiane peut correspondre à une suite tautosyllabique ou hétérosyllabique (cf. Tableau 9).

Tableau 9 : Conditions expérimentales de l'Expérience 3 et détail des groupes de consonnes utilisés.

	Tautosyllabique	**Hétérosyllabique**
Exemple ('sac')	/saklɔd/	/saktɔd/
Types de groupes	occl.-liq.	occl.-fric. occl.-occl.
Groupes consonantiques	/gl/, /kl/, /kr/	/gz/, /kt/, /ks/
Probabilités d'occurrence brute (moyenne)	74, 80, 85 (79.7)	73, 82, 85 (80)
Probabilités d'occurrence positionnelle (moyenne)	66, 70, 80 (72)	1, 1, 21 (7.7)

La fréquence des suites de consonnes médianes est contrôlée précisément à partir des données obtenues dans l'analyse distributionnelle du Chapitre 4 ; les suites tautosyllabiques ont donc, si l'on ne prend pas en compte leur position dans les mots, la même fréquence moyenne que les suites hétérosyllabiques. Six groupes de consonnes différents ont été comparés (cf. Tableau 9). Ils sont appariés du point de vue de leur probabilité d'occurrence brute mais présentent des probabilités d'occurrence positionnelles tout à fait différentes et seront donc considérés, sur la base de cet indice ainsi que sur nos intuitions de locuteur, comme présentant une structure syllabique distincte. Les suites occlusive-liquide donnent lieu à un non-alignement entre frontières lexicale et phonologique alors que les deux autres catégories de groupes produisent un alignement de ces deux types de frontières. Tout comme dans l'Expérience 1, aucune des séquences C(C)VC + C initiales ne constitue un début de mot possible dans la base de données BRULEX (Content et al., 1990). La même procédure de contrôle de ce critère a été appliquée. Un script AWK lisait un fichier contenant la liste des suites C(C)VC + C et recherchait la présence éventuelle de chaque suite en début de mot. A la fin du traitement, un fichier était créé qui contenait la liste des séquences données en entrée et le nombre de fois où chaque séquence avait été rencontrée en début de mot dans la base de données. Pour chaque séquence, on vérifiait que le nombre était bien égal à 0. Dans le cas contraire, un phonème différent était choisi et le script relancé sur les suites modifiées jusqu'à ce que l'ensemble des suites de la liste corresponde à des séquences ne constituant le début d'aucun mot dans la base de données. Vingt stimuli de remplissage ont été choisis dans lesquels aucun mot de la langue n'est enchâssé. La liste des stimuli expérimentaux et de remplissage est présentée en Annexe 14 (p.XXVI).

Les stimuli ont été enregistrés, digitalisés (16 bits, échantillonnage à 16kHz) et stockés sur le disque dur d'un ordinateur de la même manière que dans l'Expérience 1. Un exemplaire de chaque item a été sélectionné parmi trois répétitions. Les stimuli ont ensuite été édités afin de mesurer la durée entre le début du stimulus et celui de l'occlusion de la consonne finale du mot. Cette valeur est utilisée afin de mesurer la latence de la réponse à partir de la 'fin' de la portion acoustique correspondant au mot.

1.1.3. Procédure

Nous avons utilisé une tâche de word-spotting ; celle-ci a été mise en œuvre avec une procédure similaire à celle de l'Expérience 1. Les participants devaient appuyer le plus rapidement possible sur un bouton-réponse dès qu'ils détectaient un mot de la langue en position initiale de non-mot et prononcer ensuite ce mot à voix haute avant de passer à l'essai suivant. Les temps de réaction étaient mesurés depuis le début du stimulus et corrigés automatiquement par le logiciel de passation afin de refléter la latence entre l'occlusion de la consonne finale du mot à détecter et la réponse du sujet. Avant la phase expérimentale proprement dite, les participants se familiarisaient avec la tâche à l'aide d'un matériel différent.

1.2. Résultats

Nous avons utilisé les mêmes critères de sélection des participants et des items que dans l'Expérience 1. Dans la présente expérience, ces critères nous conduisent à rejeter les données d'une quantité importante de participants. Sur les quarante qui ont passé l'expérience, trente seulement ont détecté plus de la moitié des mots présents dans le matériel. Deux items ('fugue' et 'duc') ont dû être retirés de l'analyse en raison d'un taux trop important d'erreurs d'identification ou d'absence de détection de la part des 30 sujets restants. Les temps de réaction supérieurs à 1724 ms n'ont pas été introduits dans l'analyse. Les données non-intégrées dans l'analyse des latences de détection constituent 20.2 % de l'ensemble des observations.

1.2.1. Temps de réaction

Les temps de réaction et taux d'erreur observés dans l'Expérience 3 sont présentés dans le Tableau 10. Le facteur *Liste* n'interagit pas avec le facteur *Contexte*. Il n'est donc pas introduit dans les analyses de variance qui ont été réalisées sur les temps de réaction. Celles-ci sont conduites avec le *Contexte* (C_2) comme variable intra-sujet ($S * C_2$) et intra-item ($I * C_2$). La latence moyenne de détection est de 853 ms (SD = 194 ms). Les latences obtenues dans la condition de contexte hétérosyllabique sont en moyenne *plus élevées* (864 ms) que celles de la condition tautosyllabique (839 ms) ; c'est donc l'inverse de ce qui avait été observé dans l'Expérience 1. L'effet du contexte n'est néanmoins significatif ni dans l'analyse par sujet ($F_1 < 1$), ni dans l'analyse par item ($F_2 < 1$).

Tableau 10 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 3 en fonction du type de groupe consonantique.

	Tautosyllabique	**Hétérosyllabique**
TR	839 (31)	864 (40)
Taux d'erreur	16.3	24.1

Les résultats de l'analyse des temps de réaction ne vont donc pas dans le sens d'un recours à des connaissances sur les contraintes phonologiques de la langue dans les processus de segmentation lexicale. Dans cette expérience, la comparaison de groupes de consonnes de même fréquence d'occurrence dans la langue semble contribuer à faire disparaître l'effet 'phonologique' observé précédemment. Il est cependant utile d'étudier l'éventuelle émergence de cet effet dans les taux d'erreurs. Rappelons que l'effet mis en évidence par McQueen (1998) se manifeste dans les taux d'erreurs et pas dans les temps de réaction. Bien que, dans le cadre de l'analyse des résultats de l'Expérience 1, nous ayons mis en évidence un effet du contexte sur les temps de réaction, le matériel utilisé est différent. Par ailleurs, le nombre élevé de participants n'ayant pas atteint notre critère d'inclusion nous incite à penser que cette expérience est plus difficile que la première. Du fait de cette différence de difficulté, on peut envisager que l'effet attendu se manifeste dans les mesures effectuées sur une autre variable.

1.2.2. Taux d'erreurs

Les mêmes analyses ont donc été conduites sur les taux d'erreurs (20.2 % de l'ensemble des données). Le facteur *Liste* n'interagit pas avec le facteur *Contexte* ; il n'est donc pas intégré dans la présente analyse. Le sens de l'effet observé est à nouveau inversé par rapport à celui que nous avons mis en évidence dans la première expérience. On observe un pourcentage d'erreurs plus élevé dans la condition hétérosyllabique (24.1 %) que dans la condition tautosyllabique (16.3 %). Cet effet est significatif dans l'analyse par sujet ($F_1(1,29) = 6.064$, $p < .05$) et marginal dans l'analyse par item ($F_2(1,17) = 3.575$, $p = .076$). Les mots du matériel ont donc tendance à être détectés plus difficilement lorsqu'ils sont prononcés en contexte hétérosyllabique qu'en contexte tautosyllabique. La détection des mots intégrés dans les non-mots est donc ici plus rapide lorsque frontières syllabique et lexicale *ne sont pas* alignées que lorsqu'elles sont alignées.

1.3. Discussion

L'effet observé dans les taux d'erreurs est ici l'inverse de celui que nous avions mis en évidence dans l'analyse des latences de détection conduite sur les données de l'Expérience 1. Alors que dans la première expérience les mots prononcés en contexte tautosyllabique étaient détectés plus lentement que ceux de la condition hétérosyllabique, on observe ici une tendance pour les mots en condition tautosyllabique a être détectés plus facilement. Même si cet effet n'est significatif que dans l'analyse par sujet, son sens ne concorde pas avec les prédictions dérivées de l'hypothèse d'un recours à des connaissances sur les contraintes phonologiques de la langue. En effet, une segmentation lexicale fondée sur la structuration de la chaîne de parole à partir des indices phonotactiques ou de l'arborescence syllabique devrait conduire à un accroissement des taux d'erreur dans la condition de non-alignement entre frontières syllabique et lexicale. Dans cette expérience, ce sont les groupes médians tautosyllabiques qui correspondent à la condition d'alignement. Ces groupes de consonnes auraient donc dû donner lieu, selon l'hypothèse d'une segmentation lexicale fondée sur les régularités phonologiques, à des taux d'erreur plus importants que les groupes hétérosyllabiques.

Nous avons déjà observé une inversion de l'effet dans la tâche de décision lexicale de l'Expérience 2. Il a été envisagé, dans le cadre de cette expérience-contrôle, que l'inversion du sens de l'effet pouvait s'expliquer par des caractéristiques acoustiques des stimuli (par exemple la réduction consonantique décrite par Van Son & Pols, 1999), lesquelles seraient probablement déterminées par le contexte phonémique de la consonne finale du mot. Ce contexte phonémique déterminant, dans nos expériences, le caractère hétéro- ou tautosyllabique des groupes de consonnes, on observerait un lien entre la structure syllabique du groupe de consonnes médian et d'éventuelles caractéristiques acoustiques de la consonne finale du mot. Or on observe ici un effet identique à celui observé dans l'Expérience 2. On peut donc affirmer que l'on est ici en présence d'un effet lié à la structure phonologique. Mais cet effet, s'il est effectivement lié à des caractéristiques acoustiques de la consonne finale, n'est pas un effet contextuel. Il ne dépend pas de la structure phonologique d'une suite de phonèmes mais de la réalisation acoustique d'un phonème, cette réalisation acoustique étant elle seule déterminée par la structure phonologique de la séquence. Dans la présente expérience, nous avons cependant procédé à une manipulation indépendante de la structure phonologique et de la fréquence en maintenant cette dernière constante. Dans l'Expérience 1, ces deux variables étaient mélangées. On comparait donc des groupes de consonnes tautosyllabiques et hétérosyllabiques qui se distinguaient également du point de vue de leur fréquence d'occurrence dans la langue. La suppression de la composante fréquentielle qui distingue les groupes de consonnes comparés correspond ici à une disparition de l'effet d'alignement entre frontières 'prélexicale' et lexicale. Cette disparition de l'effet d'alignement lorsque les groupes de consonnes sont appariés en fréquence incite à envisager que les effets 'phonologiques' observés auparavant soient déterminés par la fréquence des groupes consonantiques plutôt que par leur rattachement à la structure syllabique ou par leur légalité phonotactique.

2. Expérience 4 : Estimation du rôle de la fréquence

Il est donc essentiel d'étudier le rôle de la fréquence des groupes de consonnes en comparant des suites dont les caractéristiques phonologiques et la syllabation sont similaires. Une telle étude nous permettrait d'isoler un effet de la fréquence. Si les effets 'phonologiques' peuvent éventuellement s'expliquer par la fréquence des groupes de consonnes comparés plus que par des connaissances abstraites sur la structure phonologique de la langue, nous devrions alors mettre en évidence un effet fréquentiel similaire à celui qui a été observé dans l'Expérience 1. Les groupes de consonnes fréquents devraient alors être regroupés à l'intérieur d'une hypothèse lexicale alors que les groupes rares auraient tendance à être considérés comme chevauchant une frontière de mots. L'occurrence d'un groupe fréquent ayant comme phonème initial la consonne finale du mot à détecter générerait alors une discordance entre frontières probabiliste et lexicale. Au contraire, ces deux frontières seraient alignées dans le cas d'un groupe consonantique rare.

Si l'effet observé dans l'Expérience 1 peut s'expliquer par une fréquence d'occurrence plus élevée pour les groupes tautosyllabiques / légaux que pour les groupes hétérosyllabiques / rares, nous devrions ici être en mesure de faire émerger un effet de l'alignement entre frontières probabiliste et lexicale dont le sens serait identique à celui observé entre groupes tautosyllabiques et hétérosyllabiques lorsque la fréquence n'est pas contrôlée. Cette hypothèse est valide quelle que soit l'origine effective de cet éventuel effet probabiliste (segmentation probabiliste prélexicale ou activation d'une quantité plus importante de candidats lexicaux). Deux paramètres peuvent évidemment contribuer à l'émergence d'un effet. Il est tout à fait possible que les données observées dans l'Expérience 1 trouvent leur source à la fois dans des connaissances sur les régularités phonologiques de la langue et dans un recours aux régularités probabilistes. Toutefois, en l'absence d'un effet d'alignement identique dans les expériences 1 et 3, l'émergence d'un effet de fréquence nous conduirait à remettre en cause la proposition d'un rôle des connaissances sur la structure phonologique de la langue dans les processus de segmentation de la parole en mots.

2.1. Méthode

2.1.1. Sujets

Vingt-trois étudiants en second cycle de psychologie à l'Université Paris 5 - René Descartes ont participé à cette expérience, soit volontairement, soit en échange de crédits d'enseignement. Ceux-ci n'avaient pas pris part aux précédentes expériences. Ils étaient tous de langue maternelle française et n'avaient jamais souffert de troubles auditifs.

2.1.2. Matériel

Les stimuli sont à nouveau des bisyllabes de structure C(C)VCCVC sans signification. La moitié d'entre eux (au nombre de 29) commence par un mot d'une syllabe correspondant à la séquence de phonèmes initiale C(C)VC (par exemple /fygzab/, 'fugue' + /zab/). Tout comme dans l'Expérience 3, nous avons été amenés à utiliser un nombre restreint de groupes consonantiques médians. Par conséquent, les mots utilisés dans le matériel ne sont pas les mêmes que ceux des précédentes expériences. Nous avons appliqué les mêmes critères de choix du matériel que dans les expériences 1 et 3. Aucun des mots à détecter ne contient une voyelle qui permettrait d'influencer le rattachement syllabique de la consonne suivante (comme /o/ ou /e/). Les mots monosyllabiques sélectionnés se terminent tous par une consonne occlusive. Chaque mot est prononcé dans deux séquences sans signification correspondant chacune à deux conditions de contexte phonémique. Ces non-mots présentent une structure C(C)VCCVC dans laquelle la suite CC médiane peut correspondre à une suite fréquente ou rare (cf. Tableau 11). La fréquence des suites de consonnes médianes est estimée à l'aide des données obtenues dans l'analyse distributionnelle du Chapitre 4. Six groupes de consonnes différents ont été comparés (cf. Tableau 11).

Tableau 11 : Conditions expérimentales de l'Expérience 4 et détail des groupes de consonnes utilisés.

	Fréquent	**Rare**
Exemple ('fugue')	/fygzab/	/fygʒab/
Types de groupes	occl.-fric. occl.-occl.	occl.-fric. occl.-occl.
Groupes consonantiques	/gz/, /ps/, /pt/	/gʒ/, /pʃ/, /pk/
Probabilités d'occurrence brute (+ moyenne)	73, 75, 72 (73.3)	44, 6, 21 (23.7)
Probabilités d'occurrence positionnelle (+ moyenne)	1, 57, 1 (19.7)	1, 1, 1 (1)

La moitié d'entre eux présente une fréquence d'apparition brute plus élevée que l'autre moitié. La valeur moyenne de probabilité positionnelle est faible pour les deux classes de groupes comparées. Ils sont appariés du point de vue de leur mode d'articulation (occlusive-fricative ou occlusive-occlusive). Ainsi, /gz/ et /gʒ/ prononcés en position intervocalique peuvent être considérés comme tout à fait comparables pour ce qui est de leur rattachement à la structure syllabique. Ils n'ont par ailleurs pas tendance à être utilisés en début de mot dans la langue. Bien que /gz/ soit à notre connaissance attesté en début de mot (c'est le cas dans 'xylophone'), ce n'est toutefois pas un groupe très prolifique dans cette position. A notre connaissance, la suite /gʒ/ n'est pas attestée en début de mot dans la langue. Ces deux groupes consonantiques sont de toutes façons très peu utilisés aux débuts de mot en français. L'analyse distributionnelle réalisée au Chapitre 4 révèle en tout cas qu'ils sont tous deux absents de la base de données BRULEX (Content et al., 1990) en position initiale de mot[42].

La probabilité d'occurrence en début de mot n'a cependant pas pu être contrôlée strictement pour les trois paires de groupes de consonnes. L'opposition /ps/ vs. /pʃ/ pose notamment un problème plus délicat. La différence de probabilité brute d'occurrence s'accompagne également d'une différence de probabilité positionnelle. Le groupe /ps/ apparaît dans un certain nombre de mots dans la base de données BRULEX (Content et al., 1990) alors que la séquence /pʃ/ -ainsi que les 4 autres groupes comparés- est totalement absente en début de mot dans ce lexique. A l'indice de probabilité positionnelle mesuré dans le Chapitre 4 correspond une fréquence d'occurrence (nombre de mots recensés) de 43. La plupart des mots de cet ensemble commencent par les morphèmes de la famille 'psych-'. Il est évident que ces morphèmes sont plutôt familiers aux étudiants qui ont participé aux expériences présentées dans cette thèse… Il est donc délicat d'affirmer, sur la base des probabilités positionnelles d'occurrence, que tous les groupes comparés sont rares en position initiale de mot. Même si /ps/ est largement attesté en position initiale de mot, il n'en reste pas moins beaucoup plus rare que la plupart des séquences occlusive-liquide rencontrées en français (pour lesquelles la valeur de probabilité positionnelle s'étend de 63 à 88) et peut à ce titre être considéré comme une suite de deux consonnes se rattachant chacune à une arborescence syllabique différente lorsqu'elles sont prononcées entre deux voyelles. A défaut de pouvoir justifier d'une similarité de structure phonologique des groupes comparés en se fondant sur ces probabilités positionnelles, il nous semble donc que ces trois paires de groupes consonantiques peuvent raisonnablement être considérées comme se syllabant toutes de manière identique en position intervocalique.

Les suites fréquentes donnent lieu à un non-alignement entre frontières lexicale et *probabiliste*. En outre, l'occurrence d'une suite fréquente donne lieu à l'activation d'un nombre élevé de candidats lexicaux. Les frontières lexicale et probabiliste sont alignées pour les groupes rares, lesquels donnent lieu à l'activation d'une plus faible quantité de candidats lexicaux. Comme dans les précédentes expériences, aucune des séquences C(C)VC + C initiales ne constitue un début de mot possible dans la base de données BRULEX (Content et al., 1990). La même procédure de contrôle de ce critère a été appliquée. Un script AWK lisait un fichier contenant la liste des suites C(C)VC + C et recherchait la présence éventuelle de chaque suite en début de mot. A la fin du traitement, un fichier était créé qui contenait la liste des séquences données en entrée et le nombre de fois où chaque séquence avait été rencontrée en début de mot dans la base de données. Pour chaque séquence, on vérifiait que le nombre était bien égal à 0. Dans le cas contraire, un phonème différent était choisi et le script relancé sur les suites modifiées jusqu'à ce que l'ensemble des suites de la liste corresponde à des séquences ne

[42] 'xylophone' y est retranscrit [ksilɔfɔn].

constituant le début d'aucun mot dans la base de données. Vingt-neuf stimuli de remplissage ont été choisis. Ils ne contiennet aucun mot de la langue. La liste des stimuli expérimentaux et de remplissage est présentée en Annexe 15 (p.XXVII).

Les stimuli ont été enregistrés, digitalisés (16 bits, échantillonnage à 16kHz) et stockés sur le disque dur d'un ordinateur de la même manière que dans les autres expériences. Un exemplaire de chaque item a été sélectionné parmi trois répétitions. Les stimuli ont ensuite été édités afin de mesurer la durée entre le début du stimulus et celui de l'occlusion de la consonne finale du mot. Cette valeur est utilisée afin de mesurer la latence de la réponse à partir de la 'fin' de la portion acoustique correspondant au mot.

2.1.3. Procédure

Les participants ont réalisé une tâche de word-spotting selon une procédure identique à celle des précédentes expériences. Ils devaient appuyer le plus rapidement possible sur un bouton-réponse dès qu'ils détectaient un mot de la langue en position initiale de non-mot puis prononcer ce mot à voix haute. Les temps de réaction étaient mesurés depuis le début du stimulus et corrigés automatiquement par le logiciel de passation afin de refléter la latence entre l'occlusion de la consonne finale du mot et la réponse du sujet. Avant la phase expérimentale proprement dite, les participants se familiarisaient avec la tâche de word-spotting.

2.2. Résultats et discussion

Nous avons utilisé les mêmes critères de sélection des participants et des items que dans les précédentes expériences. Trois participants ont détecté moins de la moitié des mots du matériel. Leurs résultats ne sont pas intégrés dans l'analyse. Deux items ('croupe' et 'ligue') ont dû être retirés de l'analyse en raison d'un taux trop important d'erreurs d'identification ou d'absence de détection de la part des 20 sujets restants. Les temps de réaction supérieurs à 1600 ms n'ont pas été introduits dans l'analyse. Les données non-intégrées dans l'analyse des latences de détection constituent 19.4 % de l'ensemble des observations.

2.2.1. Temps de réaction

Les temps de réaction et les taux d'erreur observés sont présentés dans le Tableau 12. La variable *Liste* n'interagit pas avec la variable *Fréquence* ; elle n'est donc pas introduite dans les analyses de variance qui ont été réalisées sur les temps de réaction. Celles-ci sont conduites avec la *Fréquence* (F_2) comme variable intra-sujet ($S * F_2$) et intra-item ($I * F_2$). La latence moyenne de détection est de 762 ms (SD = 160 ms). Les latences obtenues pour les groupes de consonnes fréquents et rares sont relativement similaires (respectivement 767 ms et 758 ms). L'effet de fréquence n'est significatif ni dans l'analyse par sujet ni dans l'analyse par item (tous les F sont inférieurs à 1). L'analyse des latences de détection ne permet donc pas de conclure à la présence d'un effet d'alignement entre frontières probabiliste et lexicale qui serait assimilable à l'effet d'alignement entre frontières phonologique et lexicale préalablement mis en évidence dans l'Expérience 1.

Tableau 12 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 4 en fonction de la fréquence du groupe consonantique.

	Fréquent	**Rare**
TR	767 (36)	758 (36)
Taux d'erreur	22.0	16.8

2.2.2. *Taux d'erreurs*

Les mêmes analyses ont été conduites sur les taux d'erreurs (19.4 % de l'ensemble des données). Le facteur *Liste* n'interagit pas avec le facteur *Fréquence* ; il n'est donc pas intégré dans la présente analyse. On observe un pourcentage d'erreurs plus élevé lorsque le groupe de consonnes médian est fréquent (22.0 %) que lorsqu'il est rare (16.8 %). Cet effet est significatif dans l'analyse par sujet ($F_1(1,19) = 4.930$, $p < .05$) et marginal dans l'analyse par item ($F_2(1,26) = 3.250$, $p = .083$). Les mots du matériel ont donc tendance à être détectés plus difficilement lorsqu'ils sont suivis d'une consonne qui, associée à la consonne finale du mot, donne lieu à un groupe fréquent que lorsqu'elle donne lieu à un groupe rare.

2.2.3. *Discussion*

Le sens de l'effet observé est cette fois-ci similaire à celui qui avait été mis en évidence dans les latences de détection de l'Expérience 1. Cette première expérience avait pour objet de comparer les groupes tautosyllabiques et hétérosyllabiques en contrôlant la contribution des phénomènes de sélection lexicale sérielle dans l'émergence de l'effet. Rappelons que nous avions alors observé des temps de réaction plus élevés pour les groupes tautosyllabiques que pour les groupes hétérosyllabiques. Le même type de résultats est obtenu par McQueen (1998) qui observe, lorsque les mots à détecter sont en position finale du non-mot, des taux d'erreur plus élevés pour les groupes phonotactiquement légaux que pour les groupes illégaux. Dans une tâche de détection de phonèmes, Vroomen & De Gelder (1999) avaient quant à eux observé des latences de détection plus longues pour des séquences tautosyllabiques Consonne-Voyelle que pour des suites hétérosyllabiques Consonne-Consonne. Dans ces trois expériences, les groupes tautosyllabiques / légaux correspondent toujours à des séquences fréquentes dans la langue alors que les séquences hétérosyllabiques / illégales sont relativement rares.

Si l'on assimile l'opposition *tautosyllabique vs. hétérosyllabique* (ou *légal vs. illégal*) à l'opposition *fréquent vs. rare*, nous sommes ici en présence d'un effet tout à fait similaire. L'occurrence d'une suite fréquente induisant une discordance entre frontière lexicale et frontière syllabique / phonotactique / probabiliste génère un accroissement de la difficulté à isoler le mot qui s'exprime ici par une augmentation des taux d'erreurs. Or le matériel utilisé dans cette expérience est uniquement composé de groupes consonantiques occlusive-fricative ou occlusive-occlusive qui subissent une syllabation similaire dans la chaîne de parole. Les données de cette expérience conduisent donc à envisager que l'allongement des temps de réaction observé dans l'Expérience 1 ainsi que dans les travaux de Vroomen & De Gelder (1999), de même que l'accroissement des taux d'erreur observé par McQueen (1998), lorsque frontières phonologique et lexicale ne concordent pas puisse s'expliquer -au moins en partie- par la fréquence d'occurrence des groupes de consonnes impliqués. Cet effet pourrait dépendre aussi bien de procédures de segmentation probabiliste prélexicales que sur des phénomènes de compétition entre candidats lexicaux, lesquels seraient plus nombreux lorsqu'un groupe de consonnes fréquent est produit dans une séquence de parole.

2.3. Analyses complémentaires

La marginalité de l'effet de fréquence observé dans l'analyse des taux d'erreur incite cependant à approfondir l'étude de cet effet de la même manière que dans l'analyse des données de l'Expérience 1. L'observation d'un effet seulement marginal dans l'analyse par item nous avait alors conduit à dichotomiser l'échantillon des items en deux groupes en fonction de la fréquence ou du nombre de compétiteurs des mots. Cette dichotomie avait permis de faire émerger un effet significatif aussi bien dans l'analyse par item que dans l'analyse par sujet pour les mots ayant beaucoup de compétiteurs. Une variabilité dans les caractéristiques intrinsèques des mots à détecter peut influencer leur aptitude à faire émerger un effet donné. La présence de mots pour lesquels le contexte n'a qu'un effet limité pourrait alors empêcher le test statistique de rejeter l'hypothèse nulle pour l'autre ensemble de mots. Suivant un raisonnement et un objectif identiques, le même type d'analyse a été conduit sur les données de l'Expérience 4 afin de chercher à faire émerger un effet de fréquence significatif dans l'analyse par item. Ces analyses ne sont cependant pas concluantes. Les valeurs médianes de la fréquence et du nombre de compétiteurs (respectivement 12 et 7) ont été mesurées selon la même procédure. Aucun effet de la fréquence des groupes consonantiques n'émerge pour les mots fréquents ($F_1(1,19) = 1.760$, $p > .1$; $F_2(1,13) = 2.720$, $p > .1$) pas plus que pour ceux qui ont beaucoup de compétiteurs (tous les F sont inférieurs à 1). L'effet de fréquence n'est pas non plus significatif lorsque l'analyse est conduite sur les mots les plus rares ($F_1 < 1$; $F_2(1,12) = 2.182$, $p > .1$). Cet effet est par contre significatif lorsque l'on se restreint aux items qui ont peu de compétiteurs (6 ou moins de 6 dans la base de données Brulex, Content et al., 1990). On observe alors un effet significatif aussi bien dans l'analyse par sujet ($F_1(1,18) = 11.302$, $p < .01$) que dans l'analyse par item ($F_2(1,13) = 8.381$, $p < .05$; cf. Tableau 13 pour un descriptif des taux d'erreur mesurés dans chaque liste en fonction du contexte). La dissociation en fonction du nombre de compétiteurs fait en revanche émerger des effets de liste. L'effet de la fréquence étant alors absent dans l'une des deux listes (le F est inférieur à 1 dans les analyses réalisées aussi bien avec le facteur *Sujet* qu'avec le facteur *Item* comme variable aléatoire).

Tableau 13 : Taux d'erreurs (en %) observés dans l'Expérience 4 pour les mots ayant peu de compétiteurs en fonction de la fréquence du groupe consonantique et de la liste expérimentale.

	Fréquent	**Rare**
Liste 1	35.0	18.3
Liste 2	23.3	20.0

L'émergence d'un effet de liste ne permet pas de recourir à ces résultats pour réfléchir sur l'effet de la fréquence des groupes de consonnes. Il est effectivement tout à fait possible que cet effet de fréquence observé dans l'une des deux listes soit lié à des caractéristiques spécifiques des mots qui apparaissent dans chaque contexte (fréquent vs. rare) en fonction de la liste expérimentale, les mots prononcés en contexte fréquent dans la liste 1 ayant des caractéristiques communes qui divergent de celles des mots qui sont prononcés en contexte rare dans la même liste. Il reste cependant que l'analyse conduite sur l'ensemble du matériel fait émerger un effet de la fréquence des groupes consonantiques.

2.4. Discussion générale

Les deux expériences présentées dans ce chapitre contribuent à une conception alternative des effets phonologiques observés dans les travaux antérieurs. Si l'on observe un effet d'alignement entre frontières phonologique et lexicale lorsque la fréquence des groupes de consonnes n'est pas contrôlée, la comparaison de groupes de même fréquence conduit à une disparition de l'effet. A fréquence égale, les résultats de l'Expérience 3 ne permettent donc pas de mettre en évidence un accroissement de la difficulté à localiser les frontières lexicales lorsque la consonne finale d'un mot constitue l'attaque de la syllabe suivante. Si l'effet attendu n'émerge pas lorsque l'on compare des groupes tauto- et hétéro-syllabiques de même fréquence, on observe au contraire un effet de la fréquence des groupes consonantiques dont la syllabation en position intervocalique est similaire. Les groupes fréquents donnent alors lieu à un accroissement des taux d'erreur similaire à celui qui était observé dans l'Expérience 1 pour les groupes tautosyllabiques. La combinaison des conclusions dérivées de ces deux expériences pourrait nous amener à une interprétation relativement simple : les processus de segmentation de la parole en mots ne font pas intervenir de connaissances sur les régularités phonologiques de la langue. La mise en évidence d'un effet d'alignement entre frontières phonologique et lexicale n'est qu'un artefact de la confusion entre structure phonologique et fréquence d'occurrence dans la langue. D'éventuelles connaissances sur les régularités phonologiques de la langue ne seraient donc pas utilisées par les auditeurs pour localiser les frontières de mots. Ces régularités linguistiques étant corrélatives des fréquences d'apparition dans la langue, ce sont ces différences de fréquence qui conduiraient -soit par le biais de processus prélexicaux de segmentation probabiliste, soit comme conséquences des phénomènes de compétition entre candidats lexicaux- à l'émergence d'effets assimilables à des effets phonologiques.

Il semble cependant nécessaire d'approfondir les données obtenues dans l'Expérience 3. En effet, jusqu'à maintenant les expériences que nous avons conduites portaient sur un matériel constitué à moitié d'items-test alors que l'autre moitié consistait en des items de remplissage. Cette proportion de 50 % conduit à une succession relativement fréquente d'items-test pour lesquels les participants doivent donner une réponse. Nous avons mentionné à plusieurs reprises la difficulté de la tâche de *word-spotting*. Lorsqu'une tâche est particulièrement difficile, les participants peuvent se sentir désorientés et risquent d'autant plus de chercher à mettre en place des stratégies de réponse adéquates leur permettant d'accomplir cette tâche au mieux. Ces stratégies peuvent notamment dépendre en partie des réponses qui ont été données au préalable. Ainsi, si l'on détecte un mot à l'initiale d'une séquence CVC#CVC, on peut émettre des prédictions concernant la structure syllabique de la séquence-test suivante. Si ces prédictions sont justifiées, la latence de détection peut être réduite et le risque d'erreur peut être moins important. Au contraire, si la structure syllabique du stimulus-test suivant ne concorde pas avec les attentes qui ont été développées, on risque d'observer un accroissement considérable du temps de réponse aussi bien que des erreurs plus fréquentes. Le développement d'attentes en fonction des caractéristiques des stimuli peut donc influencer les variables dépendantes et masquer l'émergence d'un effet. Plus le délai entre deux items-test est restreint, plus le risque d'influence de la réponse précédente sur la réponse en cours est élevé. Or, dans une expérience dans laquelle la moitié des stimuli correspond à un item-test, on a en moyenne un seul stimulus de remplissage entre chaque réponse. Dans l'éventualité de la mise en œuvre d'une stratégie de réponse reposant sur les réponses précédentes, le nombre d'items de remplissage est alors trop faible pour limiter l'effet de ces stratégies. Il serait donc intéressant de réduire la proportion d'items-test dans le matériel expérimental afin de modérer l'influence possible des réponses précédentes sur les attentes développées par les participants.

3. Expérience 5 : Augmentation du nombre d'items de remplissage

L'Expérience 3 a été répliquée avec une légère modification du matériel. La proportion de stimuli-test a été réduite de 50 % à moins de 25 % en augmentant le nombre de stimuli de remplissage. Cette manipulation devrait limiter le recours éventuel à des stratégies de réponse liées aux essais précédents en insérant en moyenne un nombre plus important d'essais sans réponse attendue entre chaque essai expérimental.

3.1. Méthode

3.1.1. Sujets

Vingt et un étudiants en second cycle de psychologie à l'Université Paris 5 - René Descartes ont participé à cette expérience, soit volontairement soit en échange de crédits d'enseignement. Aucun d'entre eux n'avait participé aux précédentes expériences. Ils étaient tous de langue maternelle française et n'avaient jamais souffert de troubles auditifs.

3.1.2. Matériel

Vingt items de remplissage supplémentaires ont été créés qui ne contiennent aucun mot en position initiale du non-mot bisyllabique. Le matériel de l'Expérience 5 est en grande partie identique à celui de l'Expérience 3. Les 20 items-test sont identiques et prononcés dans les deux conditions de contexte (tautosyllabique vs. hétérosyllabique). Les 20 items de remplissage originaux ont été conservés et 45 nouveaux items de remplissage ont été créés, portant le nombre total d'items de remplissage à 65. La décision de reprendre un matériel identique nous a confrontés à un dilemme. Nous avions en effet le choix entre deux solutions pour l'enregistrement des stimuli : réenregistrer la totalité du matériel ou n'enregistrer que les stimuli de remplissage supplémentaires. La première solution aurait conduit à utiliser un matériel expérimental acoustiquement différent dans les deux versions de l'expérience et aurait par ailleurs nécessité de réitérer toutes les mesures de durée déjà disponibles. La seconde solution permettait de réutiliser exactement les mêmes stimuli-test mais nous exposait au risque de différences acoustiques nettes entre les nouveaux stimuli de remplissage et les stimuli originaux (voix différente, intensité différente, niveau de bruit différent). Nous avons choisi de privilégier la seconde solution en contrôlant au maximum les conditions d'enregistrement et de digitalisation pour qu'elles soient le plus similaires possible. A cet effet, le même locuteur a réécouté la bande originale plusieurs fois avant de procéder à l'enregistrement des stimuli supplémentaires. Cette phase a permis de reprendre une voix et une intonation similaires dans des conditions techniques et acoustiques identiques (même enregistreur DAT, même cabine insonorisée, mêmes réglages, même ordinateur utilisé pour la digitalisation). Les fichiers son ont ensuite été normalisés en intensité dans l'environnement MATLAB afin d'éviter toute variation régulière de l'intensité moyenne en fonction de l'origine des stimuli. Une écoute attentive et aléatoire ne permettant pas d'identifier la provenance des différents items de remplissage, nous avons décidé de procéder à la passation expérimentale avec ce matériel composé des stimuli originaux de l'Expérience 3 et des items de remplissage supplémentaires enregistrés spécifiquement pour l'Expérience 5. Les stimuli-test sont non seulement constitués des mêmes suites de phonèmes et contiennent aussi bien les mêmes mots que les mêmes groupes de consonnes, mais ils sont en outre acoustiquement identiques dans les deux expériences (cf. Tableau 9, Section 1.1.2 pour le détail des conditions expérimentales qui sont strictement identiques à celles de l'Expérience 3).

3.1.3. Procédure

La procédure est identique à celle de l'Expérience 3. La durée de l'expérience est doublée en raison de l'accroissement important de stimuli présentés au cours de l'expérience.

3.2. Résultats

Un seul des participants a détecté moins de la moitié des mots du matériel. Ses résultats ne sont pas intégrés dans l'analyse. Seuls 13 items sur les 20 du matériel atteignent le critère de sélection. Les 7 autres items ('bac', 'bague', 'blague', 'digue', 'duc', 'fugue' et 'ligue') ont dû être retirés de l'analyse en raison d'un taux trop important d'erreurs d'identification ou d'absence de détection de la part des 20 sujets restants. Les latences de détection sont en moyenne nettement plus élevées que dans l'Expérience 3, ce qui conduit à rejeter les temps de réaction supérieurs à 2156 ms. Ces données rejetées correspondent à 23.9 % de l'ensemble des observations.

3.2.1. Temps de réaction

Les temps de réaction et les taux d'erreur observés sont présentés dans le Tableau 14. La variable *Liste* n'interagit pas avec la variable *Contexte* ; elle n'est donc pas intégrée dans les analyses de variance qui ont été effectuées sur les temps de réaction. Celles-ci sont conduites avec le *Contexte* (C_2) comme variable intra-sujet ($S * C_2$) et intra-item ($I * C_2$). La latence moyenne de détection est de 1029 ms (SD = 195 ms). Les latences obtenues pour les groupes tautosyllabiques et hétérosyllabiques sont tout à fait similaires (respectivement 1025 ms et 1034 ms). Cet effet n'est pas significatif (les F sont inférieurs à 1 aussi bien dans l'analyse par item que dans l'analyse par sujet). On retrouve ici, de même que dans l'Expérience 3, une impossibilité à mettre en évidence l'effet d'alignement entre frontières phonologique et lexicale lorsque la fréquence des groupes de consonnes médians est maintenue constante. L'accroissement du nombre d'items de remplissage supposé restreindre l'influence des stratégies de réponse sur les données mesurées ne contribue pas à faire émerger l'effet attendu dans l'analyse des latences de détection. La modification des proportions entre items de remplissage et items-test modifie cependant la difficulté de la tâche puisque latences de réponse et taux d'erreur sont en moyenne plus élevés que dans l'Expérience 3 (1029 ms vs. 853 ms ; 23.9 % vs. 20.2 %). Par ailleurs l'effet observé pourrait tout à fait, contrairement à ce qui était observé dans l'Expérience 1, se manifester dans les taux d'erreur.

Tableau 14 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 5 en fonction du type de groupe consonantique.

	Tautosyllabique	**Hétérosyllabique**
TR	1025 (37)	1034 (51)
Taux d'erreur	25.5	22.3

3.2.2. Taux d'erreur

Les mêmes analyses ont été conduites sur les taux d'erreur (23.9 % de l'ensemble des données). La variable *Liste* interagit avec la variable *Contexte* ($F_1(1,18) = 5.490$, $p < .05$). Elle est donc intégrée dans l'analyse. Les groupes tautosyllabiques donnent lieu à des taux d'erreur légèrement plus élevés (25.5 %) que les groupes hétérosyllabiques (23.3 %) ; ce qui, contrairement aux données de taux d'erreur obtenues dans l'Expérience 3, reproduit un effet d'alignement entre frontières phonologique et lexicale. Cet effet n'est cependant significatif ni dans l'analyse par item ni dans l'analyse par sujet (tous les F sont inférieurs à 1). L'interaction *Liste . Contexte* (cf. Tableau 15 pour un détail des taux d'erreur observés) se manifeste par une absence d'effet significatif du contexte dans la seconde liste expérimentale ($F_1(1,11) = 1.543$, $p > .1$). C'est par conséquent la différence de taux d'erreur observée dans la liste 1 qui est statistiquement valide. Cet effet est conforme aux prédictions dérivées de l'hypothèse d'un effet d'alignement entre frontières phonologique et lexicale, les groupes tautosyllabiques donnant lieu à des taux d'erreur plus élevés (36.7 %) que les groupes hétérosyllabiques (22.9 %). Il n'est cependant pas possible d'affirmer, du fait de sa restriction à l'une des listes, que cet effet est généré par une discordance entre frontières phonologique et lexicale plus que par des caractéristiques intrinsèques aux mots prononcés respectivement dans les conditions de contexte tauto- et hétéro-syllabique. Bien que, contrairement aux données de l'Expérience 3, le sens de l'effet soit conforme à ce que permettrait de prédire l'hypothèse d'un recours à des contraintes phonologiques indépendant de la fréquence d'occurrence des groupes de consonnes, sa restriction à une seule liste ne permet pas de confirmer l'hypothèse d'un effet d'alignement lorsque la fréquence des groupes de consonnes est maintenue constante.

Tableau 15 : Taux d'erreurs (en %) observés dans l'Expérience 5 pour les mots ayant peu de compétiteurs en fonction du type de groupe consonantique et de la liste expérimentale.

	Tautosyllabique	**Hétérosyllabique**
Liste 1	36.7	22.9
Liste 2	14.3	21.7

3.3. Discussion

L'accroissement du nombre d'items de remplissage dans le matériel ne permet pas de mettre en évidence un effet d'alignement entre frontières phonologique et lexicale. Nous avions fait l'hypothèse que cette manipulation permettrait de modérer l'influence des essais précédents sur la réponse fournie pour l'essai en cours, limitant ainsi un éventuel recours à des stratégies de réponse. Doit-on à nouveau envisager que les effets d'alignement phonologique observés au préalable se restreignent à des effets de fréquence d'occurrence des groupes consonantiques ?

Un retour aux données des expériences 3 et 4 s'impose afin de réaffirmer l'hypothèse d'un rôle des régularités phonologiques de la langue sur les processus de segmentation lexicale. On observe en effet des latences moyennes de détection tout à fait divergentes dans ces deux expériences alors que la proportion entre items-test et items de remplissage est la même. L'étude du rôle des régularités phonologiques à fréquence d'occurrence identique (Expérience 3) donne lieu à un temps de réponse moyen de 853 ms. Dans l'Expérience 4, on observe un temps de réponse moyen nettement plus rapide (762 ms) et aucun effet de fréquence sur les temps de réaction. Il est intéressant de noter que les groupes fréquents de l'Expérience 4 présentent une fréquence assez proche (73.3 sur l'échelle logarithmique des probabilités d'occurrence) de celle des deux catégories de groupes comparées dans l'Expérience 3 (respectivement 79.7 et 80). Dans l'Expérience 3, les stimuli-test présentés aux participants correspondaient à une alternance de groupes tauto- et hétéro-syllabiques. Dans l'Expérience 4 au contraire, tous les groupes comparés étaient hétérosyllabiques. On peut raisonnablement se demander pourquoi, du fait de l'absence d'effet de fréquence sur les latences de détection, les temps de réaction moyens sont à ce point plus élevés lorsque l'on mélange des groupes tauto- et hétéro-syllabiques que lorsqu'on ne présente que des groupes hétérosyllabiques. L'introduction de groupes tautosyllabiques influencerait-elle les latences de détection en rendant la tâche plus difficile ? Même si l'introduction de groupes tautosyllabiques dans le matériel correspond à un allongement important des temps de réaction, il n'est pas possible d'affirmer que cet allongement est réellement déterminé par ce paramètre puisque le matériel linguistique utilisé n'est pas identique (les mots à détecter ne sont pas les mêmes notamment). Cette observation incite cependant à réitérer notre tentative de mettre en évidence un effet du type de groupe de consonnes indépendamment de leur fréquence d'occurrence.

Nous avons jusqu'à maintenant présenté des études dans lesquelles l'apparition des stimuli correspondant aux diverses conditions expérimentales du facteur systématique principal était aléatoire. L'accroissement du nombre d'items de remplissage dans l'Expérience 5 ne change rien au fait que les stimuli pour lesquels une réponse est attendue correspondent à une alternance des conditions d'alignement et de non-alignement des frontières phonologique et lexicale. Cette alternance est voulue afin de rendre aléatoire l'occurrence de ces différentes conditions et d'empêcher les participants d'identifier une structure commune aux divers stimuli présentés. Cette alternance peut influencer les réponses des participants. En effet, si la présence d'un groupe tautosyllabique générant une discordance entre frontières phonologique et lexicale est réellement gênante pour localiser une frontière de mot, l'alternance aléatoire de stimuli provenant des deux conditions expérimentales peut perturber les réponses données. Si, conformément à ce qui a été avancé dans la discussion des données des expériences 3 et 4, les réponses fournies aux essais précédents peuvent influencer les attentes développées par les participants pour les essais suivants, une alternance des conditions pourrait perturber les données obtenues. Ce problème n'est réel que dans le cas d'une expérience particulièrement complexe dans laquelle les participants éprouvent des difficultés particulières à accomplir la tâche. Dans le cadre d'une expérience présentant un niveau de difficulté normal, cette alternance est nécessaire afin d'obtenir des données ne reposant pas sur la découverte d'une structure commune par les participants. Par contre, si l'expérience est particulièrement difficile, les problèmes engendrés par un stimulus de la condition expérimentale la plus difficile peuvent influencer les processus mis en place lors du traitement d'un stimulus moins complexe. Cette réflexion nous avait conduits à proposer une modification de la quantité d'items de remplissage afin de limiter l'influence de la réponse donnée par le sujet à l'essai suivant. Cependant, du fait de l'accroissement sensible de la difficulté de la tâche dans cette situation (7 items n'atteignent pas le critère de sélection et la latence moyenne de détection augmente d'environ 200 ms), il est possible que le nombre de stimuli de remplissage permettant de limiter l'influence des réponses d'un essai à l'autre ait augmenté simultanément, réduisant ainsi l'efficacité de cette manipulation. Il est donc toujours possible que les éventuelles difficultés engendrées par les stimuli qui correspondent à une discordance entre frontière phonologique et lexicale se répercutent sur le reste du matériel, contribuant ainsi à l'accroissement des temps de réaction ou des taux d'erreur dans les deux conditions expérimentales comparées.

4. Expérience 6 : Structure de listes semi-bloquée par condition

Afin de limiter l'effet des difficultés engendrées par l'une des conditions expérimentales sur les résultats mesurés pour l'autre, il conviendrait donc de mettre en œuvre un plan expérimental dans lequel les groupes de sujets sont bloqués par condition, chaque sujet étant soumis à une seule des deux modalités de la variable *Contexte*. Le choix d'un plan expérimental de ce type pose des problèmes d'un autre ordre puisque l'on compare alors des sujets différents pour des conditions expérimentales différentes, le facteur *Sujet* étant emboîté dans le facteur principal ($S < C_2 >$). La méthode utilisée jusqu'à présent permet au contraire d'obtenir des données de chacune des conditions expérimentales de la part de chaque sujet. Afin de combiner les avantages de ces deux plans expérimentaux ($S < C_2 >$ et $S < L_2 > * C_2$), nous avons décidé de modifier l'ordre de présentation des stimuli dans chacune des listes en présentant les stimuli groupés par condition. Chaque sujet entend alors une liste de stimuli dans laquelle apparaissent aussi bien des stimuli de la condition alignée que de la condition non-alignée. Cependant, chacun entend d'abord la moitié des stimuli dans une condition (par exemple 'alignée') puis l'autre moitié dans l'autre condition (par exemple 'non-alignée'). Cette procédure devrait limiter la propagation d'une éventuelle difficulté introduite par la condition de non-alignement au traitement des stimuli de la condition alignée.

4.1. Méthode

4.1.1. Sujets

Vingt-huit étudiants en second cycle de psychologie à l'Université Paris 5 - René Descartes ont participé à cette expérience, soit volontairement soit en échange de crédits d'enseignement. Aucun d'entre eux n'avait participé aux précédentes expériences. Ils étaient tous de langue maternelle française et n'avaient jamais souffert de troubles auditifs.

4.1.2. Matériel

Les stimuli de l'Expérience 3 ont été réutilisés sans modification. Le matériel est constitué pour moitié d'items-cible (20 mots prononcés chacun dans les deux conditions de contexte phonologique). L'autre moitié des stimuli présentés correspond aux 20 stimuli de remplissage utilisés dans la première version de l'Expérience 3.

4.1.3. Procédure

Alors que dans les expériences 3 et 5 chaque participant avait à traiter les stimuli de chaque condition dans un ordre aléatoire, la procédure de l'Expérience 6 consiste à présenter à chaque participant la moitié des mots dans une condition (par exemple les stimuli pour lesquels frontières phonologique et lexicale sont discordantes) puis l'autre moitié dans l'autre condition (ici la condition de concordance des frontières). Des items de remplissage sont intercalés entre les items-test. Le facteur secondaire *Liste* est donc conservé. Chaque sujet n'entend un même mot que dans l'une des deux conditions expérimentales. Les stimuli sont cependant regroupés en début ou en fin de liste en fonction de la condition d'alignement (c'est à dire de la catégorie du groupe consonantique médian) à laquelle ils correspondent. Un second facteur secondaire est donc introduit afin de vérifier l'absence d'un éventuel effet d'*Ordre* de passation (tauto- *puis* hétéro-syllabique vs. hétéro- *puis* tauto-syllabique). A l'exception de cette modification de la structure des listes, la procédure expérimentale est strictement identique à celle de l'Expérience 3.

4.2. Résultats et discussion

Les critères de sélection des données sont les mêmes que dans les précédentes expériences. Huit participants ont détecté moins de la moitié des mots du matériel. Leurs résultats ne sont pas intégrés dans l'analyse. Trois items ('fugue', 'fougue', 'nuque') ont dû être retirés de l'analyse en raison d'un taux trop important d'erreurs d'identification ou d'absence de détection de la part des 20 sujets restants. Les temps de réaction supérieurs à 1740 ms n'ont pas été introduits dans l'analyse. Les données non-intégrées dans l'analyse des latences de détection constituent 22.4 % de l'ensemble des observations.

4.2.1. Temps de réaction

Les temps de réaction et les taux d'erreur observés sont présentés dans le Tableau 16. La variable *Liste* n'interagit que de façon marginale avec la variable *Contexte* ($F_1(1,18) = 3.114$, $p = .095$). Elle n'est pas intégrée dans les analyses de variance qui ont été conduites sur les latences de détection. Aucun autre effet d'interaction n'est significatif. Les analyses ont été réalisées avec le *Contexte* (C_2) comme variable intra-sujet ($S * C_2$) et intra-item ($I * C_2$). Les latences de détections sont en moyenne plus longues pour les stimuli contenant un groupe médian tautosyllabique (875 ms) que pour ceux qui contiennent un groupe hétérosyllabique (811 ms). Cet effet est significatif dans l'analyse par sujet ($F_1(1,19) = 6.022$, $p < .05$) mais seulement marginal dans l'analyse par item ($F_2(1,16) = 4.216$, $p = .057$).

Tableau 16 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 4 en fonction de la structure du groupe consonantique.

	Tautosyllabique	**Hétérosyllabique**
TR	875 (45)	811 (43)
Taux d'erreur	24.7	20.3

On observe donc dans cette expérience l'effet d'alignement attendu. Les latences de réponse sont plus longues lorsque le groupe consonantique médian est prononcé à l'attaque de la seconde syllabe du stimulus que lorsqu'il chevauche une frontière syllabique. Cet effet émerge malgré la stricte équivalence des groupes comparés sur l'échelle des probabilités d'occurrence. Des groupes de consonne peuvent donc, en introduisant une discordance entre frontières phonologique et lexicale, ralentir la localisation des frontières de mot par rapport à des groupes générant un alignement de ces deux frontières. Cet effet est indépendant de procédures de segmentation qui seraient fondées sur la fréquence des suites de phonèmes. On notera cependant que, tout comme dans les données de l'Expérience 1, cet effet n'atteint le seuil statistique de significativité que dans l'analyse par sujet. Avant de procéder à des analyses complémentaires, il convient cependant de présenter les résultats de l'analyse des taux d'erreur.

4.2.2. Taux d'erreur

Les mêmes analyses ont été conduites sur les taux d'erreur (22.3 % de l'ensemble des données). Le facteur Liste interagit avec le facteur Contexte ($F_1(1,18) = 16.656$, $p < .01$). Aucune autre interaction n'émerge de l'analyse. Malgré des taux d'erreur légèrement plus élevés pour les groupes médians tautosyllabiques (24.7 %) que pour les groupes hétérosyllabiques (20.3 %), cette différence n'est pas significative ($F_1(1,18) = 1.161$, $p > .1$; $F_2(1,16) = 2.133$, $p > .1$).

Tout comme dans l'Expérience 5, l'interaction *Liste.Contexte* se manifeste par une absence d'effet dans la seconde liste. Il n'y a cependant pas de lien à voir entre ces deux observations. En effet, les deux listes de chacune des expériences dans lesquelles une procédure différente a été utilisée (alternance aléatoire vs. listes semi-bloquées) n'étaient pas constituées des mêmes associations mot - contexte. Dans l'Expérience 6, l'ordre d'apparition des *mots* est identique à celui des autres expériences. Les listes semi-bloquées ont été générées en attribuant une condition expérimentale donnée aux 10 premiers mots de la liste et l'autre condition aux 10 derniers. Ce ne sont donc pas les mêmes stimuli qui constituent les listes 1 et 2 de l'Expérience 6 et les listes 1 et 2 des expériences 3 et 5. Il n'est pas facile d'interpréter cet effet de liste, pas plus que dans les autres expériences. Il peut être lié à des caractéristiques acoustiques des stimuli mais il pourrait aussi s'expliquer par la faible quantité de mots utilisés dans le matériel des expériences qui sont présentées ici. Un léger déséquilibre du nombre d'erreurs peut alors engendrer des différences importantes de pourcentage. Tout ce qu'il est possible d'affirmer ici est que l'on n'observe pas d'effet de l'alignement entre frontières phonologique et lexicale sur les proportions d'erreur commises par les participants. Il est par contre essentiel d'approfondir l'analyse de cet effet sur les latences de détection.

4.2.3. Analyses complémentaires

Nous observons à nouveau un effet dont la significativité statistique ne se manifeste que dans l'analyse par sujet. Il n'est que marginal dans l'analyse par item. Nous avons été confronté plusieurs fois à cette difficulté dans les expériences précédentes. Des analyses complémentaires ont par conséquent été conduites sur les données de l'Expérience 6 en dissociant les mots de l'échantillon sur la base de leur fréquence ou de leur nombre de compétiteurs. Elles ont été réalisées selon la même procédure que dans l'analyse des résultats des expériences 1 et 4. Les valeurs médianes de la fréquence et du nombre de compétiteurs des mots du matériel ont été calculées afin de séparer, pour chaque analyse, l'échantillon des mots en deux groupes de taille similaire. L'analyse des latences de détection a été répliquée pour chacune des catégories de mots.

4.2.3.1. Contexte phonologique et fréquence d'usage

La médiane de la fréquence d'usage des mots est égale à 12. L'analyse a été conduite d'une part sur les mots qui ont une valeur de fréquence d'usage strictement supérieure à 12 sur un million (N = 9) et, d'autre part, sur les mots dont la fréquence sur un million est inférieure ou égale à 12 (N =8).

Pour les mots les plus fréquents, la latence moyenne de détection est de 789 ms (SD = 199 ms). Ni le facteur *Liste* ni le facteur *Ordre* n'interagissent avec le type de contexte phonologique. (tous les F sont inférieurs à 1). On n'observe pas non plus d'interaction de second ordre ($F_1 < 1$; $F_2 < 1$). Seule la variable Contexte est donc intégrée dans les analyses présentées. Les mots sont détectés plus lentement lorsque le groupe médian est tautosyllabique (833 ms) que lorsque ce groupe est hétérosyllabique (746 ms). Cet effet est significatif dans l'analyse par sujet ($F_1(1,19) = 8.134$, $p < .05$) mais aussi dans l'analyse par item ($F_2(1,8) = 6.130$, $p < .05$).

Les latences de détection observées sur l'échantillon des mots les plus rares sont nettement plus élevées (889 ms, SD = 273 ms). Le facteur *Liste* interagit de façon marginale avec le facteur *Contexte* ($F_1(1,18) = 4.141$, $p = .057$). Aucun autre effet d'interaction ne présente une tendance à la significativité (tous les F sont inférieurs à 1). Les latences de détection observées en fonction du type de groupe consonantique médian divergent peu (respectivement 876 ms et 901 ms pour les groupes tautosyllabiques et hétérosyllabiques). Cette différence n'est pas significative ($F_1 < 1$; $F_2 < 1$).

4.2.3.2. Contexte phonologique et nombre de compétiteurs

Dans l'Expérience 1, c'est l'analyse en fonction du nombre de compétiteurs des mots qui avait permis de faire émerger au mieux un effet stable du contexte phonologique. On peut donc s'attendre à mettre également en évidence un effet du contexte sur les mots de l'Expérience 6 qui ont le plus de compétiteurs. La valeur médiane du nombre de compétiteurs des mots de l'échantillon est égale à 7. Les analyse de variance complémentaires ont été conduites d'une part sur les mots qui ont plus de 7 compétiteurs dans la base de données BRULEX (Content et al., 1990). Ceux-ci sont au nombre de 9. La seconde analyse est réalisée sur les mots qui ont 7 compétiteurs ou moins (N =8).

L'analyse effectuée sur les mots qui ont beaucoup de compétiteurs (TR moyen 836 ms, SD = 205 ms) fait émerger plusieurs effets d'interaction avec le facteur *Liste*. L'interaction de second ordre est significative ($F_1(1,16) = 4.936$, $p < .05$) ; de même que l'interaction *Liste* - *Contexte* ($F_1(1,16) = 14.493$, $p < .01$). Les autres effets d'interaction n'atteignent pas le seuil de significativité (tous les F sont inférieurs à 1). Malgré ces interactions, l'effet de contexte est significatif dans l'analyse par sujet ($F_1(1,16) = 12.520$, $p < .01$) ainsi que dans l'analyse par item ($F_2(1,8) = 5.955$, $p < .05$). Les latences de détection moyennées sur les autres conditions sont plus élevées pour les groupes tautosyllabiques (889 ms) que pour les groupes hétérosyllabiques (783 ms).

Pour les mots qui ont peu de compétiteurs, la latence moyenne de détection est de 834 ms (SD =224 ms). Le facteur Liste n'interagit avec aucun des autres facteurs. L'interaction *Ordre* - *Contexte* est significative dans l'analyse par sujet ($F_1(1,18) = 4.615$, $p < .05$). Elle n'atteint pas le seuil de significativité dans l'analyse par item ($F_2 < 1$). Quoi qu'il en soit, l'effet de contexte -tout comme dans l'analyse restreinte aux mots les plus rares- n'atteint pas le seuil de significativité ($F_1 < 1$; $F_2 < 1$) ; les latences de détection étant similaires pour les groupes tautosyllabiques (876 ms) et hétérosyllabiques (901 ms).

La restriction de l'analyse aux mots qui ont le plus de compétiteurs permet donc également de faire émerger l'effet de contexte attendu dans les deux types d'analyse. Cette restriction conduit cependant à faire émerger des effets d'interaction difficiles à interpréter. Il convient donc de ne pas tenir compte de cet effet pour l'interprétation des données. En effet, il est probable qu'il ne se manifeste que dans l'une des conditions expérimentales, le moyennage sur l'ensemble des conditions contribuant malgré tout à le faire émerger en raison de l'importance de la différence dans l'une de ces conditions. Il est par conséquent possible que l'émergence de cet effet soit lié au hasard (du fait de la faible quantité d'observations par condition) ou à des particularités acoustiques des items prononcés dans chaque condition du croisement des facteurs *Contexte*, *Ordre* et *Liste*.

4.3. Discussion générale

Malgré les résultats négatifs obtenus dans les expériences 3 et 5 et la mise en évidence d'un effet de fréquence assimilable à l'effet phonologique dans l'expérience 4, les données de l'expérience 6 contribuent à confirmer le rôle de connaissances sur les régularités phonologiques de la langue dans les processus de segmentation lexicale. Conformément aux propositions de McQueen (1998) et de Vroomen & De Gelder (1999), il semble que le système cognitif effectue une segmentation du signal de parole qui serait fondée sur les régularités phonologiques de la langue. Ces procédures de segmentation contribueraient à influencer le découpage de la chaîne de parole en mots. Lorsque frontières phonologique et lexicale sont alignées, la localisation d'une frontière de mot serait plus facile, l'occurrence d'une frontière phonologique (syllabique ou phonotactique) favorisant l'hypothèse d'une frontière de mot entre les deux phonèmes considérés. Lorsque ces frontières ne sont pas alignées au contraire, la segmentation syllabique / phonotactique ne correspondrait pas au découpage lexical et l'utilisation de procédures prélexicales de segmentation fondées sur les régularités phonologiques de la langue conduirait à un découpage des groupes de phonèmes qui ne faciliterait pas -ou rendrait plus difficile- la localisation d'une frontière lexicale.

Lors de la discussion consacrée aux données de l'expérience 1, la restriction de l'effet phonologique à un sous-ensemble des mots de l'échantillon avait été considérée comme un indice du lien étroit qui pouvait exister entre structure phonologique des groupes de consonnes et nombre de mots contenant cette séquence. Il avait alors été proposé que cette restriction pouvait être le signe d'un effet d'alignement qui trouverait sa source dans des procédures lexicales plutôt que prélexicales. Un maintien constant de la fréquence des groupes de consonnes n'empêche cependant pas cet effet de se manifester dans l'expérience 6 ; ce qui infirme l'hypothèse d'une source lexicale liée au nombre de mots qui entrent en compétition. Puisque les groupes de consonnes sont de même fréquence, leur traitement donne lieu à l'activation d'un nombre similaire de candidats lexicaux (ils apparaissent dans une quantité de mots comparable). Pourquoi cet effet se restreint-il alors à nouveau à une partie du matériel ? Si l'effet est réellement prélexical, comment peut-il à nouveau dépendre de paramètres lexicaux ? Plusieurs explications peuvent être proposées. Il est possible que le recours à des connaissances phonologiques intervienne au cours de l'étape décisionnelle ; l'utilisation de ces connaissances dépendrait alors des caractéristiques des mots qui ont été activés. La question se pose alors de savoir quelles caractéristiques des mots les rendent plus faciles ou plus difficiles à identifier. Dans la première expérience, l'effet émerge pour les mots qui ont beaucoup de compétiteurs ; le même effet émerge pour les mots fréquents dans l'expérience 6. On peut penser que les mots qui ont beaucoup de compétiteurs seront particulièrement difficiles à identifier parce qu'ils entrent en compétition avec une quantité importante de candidats. Les mots fréquents sont au contraire relativement faciles à identifier ; ils donnent lieu à des latences de détection plus courtes que les mots rares dans une tâche de décision lexicale. Il semble donc délicat de proposer une interprétation décisionnelle de cette restriction. D'autre part, ces caractéristiques lexicales pourraient également correspondre à des régularités sublexicales que nous n'avons pas identifiées. La restriction à un sous-ensemble des mots dépendrait alors de critères prélexicaux tels que la fréquence des diphones. Il serait alors possible de concevoir cet effet comme un effet réellement prélexical qui entre en interaction avec une autre variable prélexicale ; ce qui ne poserait par conséquent pas de problème d'interprétation. Cette question peut-être insérée dans une problématique plus large concernant le mode d'intégration des processus prélexicaux et lexicaux dans un modèle du traitement de la parole. Celle-ci sera abordée, à titre de réflexion sur un modèle des processus cognitifs impliqués, dans la section 2.2 du Chapitre 7. L'objet du Chapitre 7 est de fournir une réflexion, assortie de quelques données expérimentales, sur le niveau d'implémentation de ces procédures de segmentation syllabique / phonotactique et sur leur mode d'intégration, c'est à dire sur leur rapport aux autres processus impliqués.

Résumé

Les expériences présentées dans ce chapitre avaient pour objet de dissocier les contributions respectives de la fréquence et des régularités phonologiques dans l'émergence de l'effet d'alignement observé dans les précédentes études. Malgré l'absence d'effet de la structure phonotactique sur les latences de détection dans les expériences 3 et 5, il est possible -dans des conditions particulières de procédure expérimentale et pour un sous-ensemble des mots- d'observer un effet d'alignement entre frontières phonologique et lexicale lorsque la fréquence d'occurrence des groupes de consonnes comparés est maintenue constante. Ces données confirment donc en partie les propositions de McQueen (1998) et de Vroomen & De Gelder (1999). La restriction de l'effet à un sous-ensemble des mots ainsi qu'à une procédure expérimentale spécifique incite cependant à développer une réflexion sur les fondements cognitifs de cet effet.

Chapitre 7

Niveau d'implémentation, Mode d'intégration

Ce chapitre a fait l'objet de communications affichées dans des congrès internationaux avec comité de lecture assortis d'une publication dans des actes :

JIOSC'97, Journées Internationales d'Orsay sur les Sciences Cognitives, 1-2 Décembre 1997, Orsay, France.

Forum Acusticum'99, Joint 137th meeting of the Acoustical Society of America and 2nd convention of the European Acoustics Association, 14-19 Mars 1999, Berlin, Allemagne.

NIVEAU D'IMPLEMENTATION, MODE D'INTEGRATION

L'objet de ce travail était d'approfondir l'étude du rôle éventuel de connaissances sur les régularités phonologiques de la langue dans les processus de segmentation de la parole en mots. La nécessité d'un contrôle plus précis du matériel linguistique utilisé a été mise en évidence dans le cadre d'une analyse de la fréquence des groupes de consonnes dans un lexique informatisé de la langue française. Cette étude, présentée dans le Chapitre 4, permet de mettre en évidence un lien étroit entre la structure syllabique des groupes de deux consonnes et leur fréquence d'occurrence dans les mots de la langue. Les groupes consonantiques tautosyllabiques s'avèrent en général plus fréquents que les groupes hétérosyllabiques. Cette observation conduit à envisager diverses interprétations possibles des effets mis en évidence par McQueen (1998) et par Vroomen & De Gelder (1999). Trois axes interprétatifs ont été proposés à partir de cette observation. Le premier se rapporte au modèle COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987). Un déséquilibre de la fréquence des séquences de phonèmes comparées risque d'engendrer également un déséquilibre du nombre de candidats lexicaux qui sont maintenus dans la cohorte lorsque le phonème contextuel a été identifié. Cette interprétation a été mise à l'épreuve des données expérimentales dans les expériences 1 et 2 (Chapitre 5). Les deux autres interprétations proposées sont directement liées à la fréquence des séquences de phonèmes. Deux classes de modèles peuvent en effet rendre compte des effets observés si, dans le matériel utilisé, structure syllabique / phonotactique et fréquence d'occurrence sont liées. Dans le cadre d'un modèle probabiliste de la segmentation lexicale, les suites fréquentes ont tendance à être regroupées à l'intérieur d'une hypothèse lexicale alors que les suites rares favorisent l'insertion d'une frontière de mots. Si ce modèle nécessite le recours à une conception prélexicale des processus de segmentation, il n'est nullement nécessaire de supposer que les informations utilisées par le système cognitif pour savoir si une suite est rare ou fréquente sont d'ordre phonologique. Le recouvrement des distributions de probabilité d'occurrence décrit dans la section 2.2.2 du Chapitre 4 confirme cette indépendance des représentations phonologiques et probabilistes. Par ailleurs, la fréquence des suites de phonèmes détermine également le nombre de candidats lexicaux qui seront activés dans le cadre d'un modèle à alignement exhaustif (McClelland & Elman, 1986 ; Norris, 1994). Cette classe de modèles peut donc prédire les résultats observés, la sélection du candidat approprié se faisant parmi une quantité de compétiteurs plus importante lorsqu'une séquence de phonèmes est fréquente que lorsqu'elle est rare. L'effet observé serait alors, en réalité, localisé à un niveau lexical de traitement. Le recouvrement des distributions observé dans le Chapitre 4 fournit la possibilité de dissocier structure phonologique et fréquence pour la comparaison de diverses catégories de groupes consonantiques. Ce second groupe d'interprétations a été testé dans les expériences 3 à 6 (Chapitre 6).

1. Résumé des résultats obtenus

Les expériences présentées dans les chapitres 5 et 6 ont été conduites afin de tester l'hypothèse d'un rôle des contraintes phonologiques dans les processus de segmentation lexicale. Les caractéristiques du matériel expérimental ont été contrôlées afin de départager les diverses interprétations proposées.

1.1. Effets observés

1.1.1. Contraintes phonologiques et sélection lexicale

Les expériences du Chapitre 4 (notamment l'Expérience 1) permettent d'affirmer que, même en l'absence d'un déséquilibre dans la taille de la cohorte de candidats, il est possible de mettre en évidence un effet de l'alignement entre frontière phonologique / probabiliste et frontière lexicale. Les latences de détection observées dans la tâche de word-spotting sont plus longues lorsque l'occlusive finale du mot est suivie d'une liquide que lorsque elle est suivie d'une fricative ou d'une occlusive. Cet effet se restreint cependant à une partie du matériel ; en effet l'analyse par item ne s'avère significative que si l'on se restreint aux mots de l'échantillon qui ont le nombre de candidats le plus élevé. Un critère lexical détermine donc l'intervention de processus censés intervenir au cours des étapes prélexicales du traitement. Cette incongruité conduit à accentuer l'idée d'une interprétation lexicale de l'effet.

1.1.2. Régularités phonologiques, fréquence et compétitions lexicales

Les expériences présentées dans le Chapitre 6 ont été conduites afin de dissocier spécifiquement les contributions de la structure phonologique et de la fréquence -celle-ci déterminant également la contribution des processus de compétitions lexicales dans l'émergence des effets observés. Dans l'Expérience 3, des groupes de consonnes tautosyllabiques (occlusive-liquide) et hétérosyllabiques (occlusive-fricative et occlusive-occlusive) de même fréquence ont été comparés. On n'observe alors aucun effet de la structure phonologique. Au contraire, l'analyse des taux d'erreur fait émerger un effet de fréquence lorsque l'on compare des groupes de consonnes qui sont tous hétérosyllabiques. Cet effet n'est cependant significatif que dans l'analyse par sujet.

L'expérience 3 a été répliquée à deux occasions en modifiant le nombre d'items de remplissage (expérience 5) puis la structure des listes expérimentales (expérience 6). Seule cette dernière expérience permet de faire émerger un effet phonologique indépendant de la fréquence. Celui-ci se manifeste à nouveau par des latences de détection plus élevées lorsque le groupe consonantique médian donne lieu à une discordance entre frontières syllabique / phonotactique et lexicale. Il se restreint à nouveau à une partie du matériel utilisé ; seule une dichotomie fondée sur la fréquence permet, pour les mots les plus fréquents, de faire émerger cet effet dans les deux analyses sans conduire à une interaction avec les variables secondaires.

1.2. Discussion générale

Les données expérimentales permettent donc, au moins en partie, de conclure à un rôle des connaissances sur les régularités phonologiques de la langue dans les processus de segmentation de la parole en mots. Plusieurs problèmes doivent cependant être posés afin de mieux comprendre comment des représentations phonologiques pourraient influencer le traitement du signal de parole. Ce dernier chapitre a pour objet de discuter du niveau d'intervention de ces processus et de leur mode d'intégration dans un modèle du fonctionnement cognitif. Deux expériences sont présentées. La suite du chapitre constitue une réflexion sur les résultats obtenus dans ce travail de thèse.

1.2.1. Processus de traitement précoces ou stratégies ?

L'un des problèmes posé par les études présentées est l'absence pure et simple d'effets dans certaines situations expérimentales. Ainsi, lorsque l'on supprime la différence de fréquence entre les groupes tautosyllabiques et hétérosyllabiques comparés, la succession aléatoire d'essais de l'une et l'autre condition semble ne pas permettre l'émergence de l'effet d'alignement (expériences 3 et 5). Cet effet émerge par contre, avec le même matériel expérimental, lorsque les conditions de concordance et de discordance des frontières sont regroupées en début et en fin de liste. Si l'émergence de cet effet dépend d'un changement de procédure, il est judicieux de poser la question de la validité de ces processus dans le cadre des procédures naturelles de traitement de la parole ; ce qui conduit à se demander si la mise en évidence de cet effet n'est pas déterminée par la difficulté de la tâche de *word-spotting* ; ces processus de segmentation phonologique n'étant utilisés que pour accomplir au mieux la tâche dévolue aux participants. Il serait alors périlleux d'affirmer que les locuteurs adultes ont effectivement recours à des connaissances sur les régularités phonologiques de la langue pour segmenter le signal de parole en mots si ces procédures ne sont utilisées que dans le cadre de la situation expérimentale ! Ces procédures devraient ainsi être considérées comme relevant de procédures stratégiques et intervenant par conséquent au cours de l'étape décisionnelle.

Nous avons déjà abordé la problématique du rôle des stratégies de réponse lors des discussions concernant les expériences 3 et 5 et avons envisagé que le développement d'attentes ait pu masquer l'émergence de l'effet attendu. Les stratégies de réponse développées par les participants constituent typiquement une classe de processus décisionnels, lesquels ne sont mis en œuvre qu'une fois l'intégralité des traitements effectuée. L'expérience 6, dans laquelle un effet significatif du contexte phonologique est mis en évidence pour les mots les plus fréquents de l'échantillon, a été mise en place afin de limiter l'influence de stratégies de réponse sur les mesures effectuées. La mise en évidence de l'effet dans cette expérience devrait donc contribuer à invalider l'idée d'un effet stratégique. Il est cependant possible que l'utilisation même de procédures de segmentation syllabiques / phonotactiques soit déterminée par un niveau optimal de difficulté de la tâche. Ainsi, lorsque la tâche est trop difficile -ceci semble notamment être le cas lorsque la contribution de la fréquence est neutralisée et que l'ordre de présentation des stimuli de chaque condition est aléatoire- les réponses des participants seraient perturbées ; ce qui conduirait à masquer l'effet phonologique. Lorsque la tâche des participants est facilitée par un regroupement des conditions de contexte en début et en fin de liste, cette succession séquentielle des conditions expérimentales limiterait l'influence des réponses précédentes sur la réponse en cours et favoriserait l'émergence de l'effet. Le paradigme de word-spotting reste néanmoins particulièrement difficile quelle que soit la procédure employée -preuve en sont les taux d'erreurs élevés dans toutes les expériences qui sont présentées ici. Il est légitime de se demander si cet effet se maintiendrait dans une tâche plus facile. Si l'utilisation de procédures de segmentation phonologique n'est mise en place que pour compenser une difficulté particulière à isoler les mots dans le matériel, l'effet disparaîtrait probablement. Si au contraire, ces procédures sont réellement mises en œuvre par le système cognitif lorsqu'il traite un signal de parole -que ce soit dans une situation naturelle ou expérimentale- l'effet devrait persister avec une tâche plus facile.

1.2.2. Lien entre régularités phonologiques et propriétés lexicales

L'autre difficulté à laquelle nous sommes confronté porte sur les résultats des analyses de variance dans lesquelles le facteur aléatoire est le facteur *Item*. Aucune des expériences présentées ne permet de faire émerger un effet clair sur l'ensemble du matériel. Seule une restriction des analyses de variance à un sous-ensemble des mots permet de faire émerger un effet significatif de l'alignement entre frontières phonologique et lexicale dans l'analyse par item. Il semble donc clair que le recours à des connaissances sur les contraintes phonologiques de la langue n'est effectif -en tout cas dans le cadre des expériences présentées ici- que pour les traitements réalisés sur certains mots du lexique. Cette restriction déterminée par une dichotomie qui est fondée sur un critère lexical pose le problème de l'interprétation à donner de l'effet observé. Si la prise en compte d'informations prélexicales fondées sur les caractéristiques des groupes de consonnes médians n'est possible qu'en fonction de certains paramètres lexicaux, cette information censée être utilisée au cours d'étapes précoces du traitement doit être envisagée différemment. En effet, les informations prélexicales sont supposées intervenir avant même la mise en place des processus lexicaux. Pour que l'effet d'une variable prélexicale puisse dépendre de paramètres lexicaux, il faut que cette variable soit en réalité prise en compte à une étape ultérieure. Cette étape peut être envisagée selon deux angles. Ces deux possibilités découlent de l'interprétation que l'on donne du nombre de compétiteurs d'un mot.

L'effet du nombre de compétiteurs d'un mot peut être implémenté à deux niveaux de représentation. Vitevitch, Luce, Pisoni, & Auer (1999 ; cf. aussi Luce, 1998 ; Vitevitch & Luce, 1999 pour une présentation du modèle NAM[43] sur lequel est fondé ce travail) présentent un modèle de l'implémentation des paramètres liés au nombre de voisins phonologiques d'un mot. Au niveau lexical, un mot ayant beaucoup de voisins phonologiques, c'est à dire pour lequel il existe beaucoup d'autres mots dans le lexique qui partagent 1 à N phonèmes avec ce mot, entrera en compétition avec un nombre élevé de candidats lexicaux. Un accroissement du nombre de voisins lexicaux rendra sa reconnaissance plus difficile puisqu'il faut alors sélectionner le candidat correct parmi une quantité plus élevée de mots possibles. Au niveau prélexical, on peut prédire un effet inverse. Si un mot a beaucoup de voisins lexicaux, les diphones dont il est constitué sont nombreux (puisque présents dans une quantité importante de mots). Cette fréquence pourra alors faciliter l'identification des phonèmes qui le constituent (cf. le rôle des probabilités transitionnelles et leur rapport avec le paradigme de la *théorie de l'information*, section 2.1.3 du Chapitre 4) et, de fait, l'identification des mots. Observer une interaction entre des variables respectivement prélexicale et lexicale peut donc s'expliquer en réalité par une interaction faisant intervenir deux variables prélexicales… ou deux variables lexicales. En ce qui concerne la restriction de l'effet du contexte phonologique à un ensemble de mots partageant des caractéristiques lexicales (le nombre de compétiteurs), on peut donc imaginer que ce que nous avons considéré jusqu'à maintenant comme une variable lexicale désigne peut-être aussi une spécificité des mots de notre échantillon qui pourrait en fait s'exprimer en termes sublexicaux. Il est donc intéressant d'analyser plus en détails les caractéristiques du matériel utilisé pour interpréter la restriction de l'effet à un sous-ensemble des mots de l'échantillon.

1.2.3. Mode d'intégration des processus de segmentation

La dernière problématique que nous souhaitons aborder dans cette thèse concerne la manière dont ces processus de segmentation phonologique entrent en relation avec les autres processus de traitement impliqués dans la perception de la parole. Plus précisément, il semble essentiel d'analyser les rapports qui s'établissent entre ces processus de segmentation phonologique supposés prélexicaux et le niveau de traitement lexical. Selon Norris, McQueen, Cutler, & Butterfield (1997), le découpage de la chaîne de parole à partir de connaissances sur les régularités de la langue servirait à moduler les niveaux d'activation lexicale. Une analyse approfondie de cette proposition aboutit à la conclusion que l'effet est, en dernier recours, lié à des différences de durée nécessaire pour qu'une unité lexicale atteigne un seuil d'activation. Cette proposition sera détaillée dans la section 3.1.1. Nous proposerons alors une interprétation alternative dans le cadre de laquelle les notions de *processus attentionnel* et de *groupement*

[43] *Neighbourhood Activation Model.*

perceptif sont primordiales. L'objet de la dernière section de ce chapitre est donc de présenter et de justifier cette hypothèse en proposant une méthode d'investigation adéquate.

2. Niveau d'implémentation

L'effet d'alignement entre frontières phonologique et lexicale mis en évidence par McQueen (1998) et par Vroomen & De Gelder (1999) peut être observé malgré les divers contrôles qui ont été effectués sur le matériel et, notamment, sur la fréquence des groupes de consonnes comparés. Cet effet n'apparaît cependant pas dans toutes les expériences réalisées. Ceci conduit à poser la question de la pertinence des processus de segmentation pour le traitement de la parole en situation naturelle ; ce qui revient à soulever le problème du caractère stratégique ou décisionnel de ces procédures de découpage phonotactique ou syllabique dans la tâche de *word-spotting*. Lorsqu'il émerge, l'effet se restreint à un sous ensemble des mots. Le lien entre l'émergence de l'effet phonologique et des paramètres lexicaux comme la fréquence ou le nombre de compétiteurs incite à approfondir l'étude de la restriction de l'effet à un sous-ensemble du matériel lexical en analysant la relation qui peut exister entre des paramètres lexicaux (par exemple la densité de voisinage lexical) et des paramètres sublexicaux (comme la fréquence des diphones qui les constituent).

2.1. Traitement de l'information ou stratégies post-perceptives ?

S'il est possible d'obtenir un effet de l'alignement entre frontières phonologique et lexicale, cet effet n'émerge que dans des conditions particulières de procédure expérimentale. Ces restrictions doivent être analysées en détail afin de mieux comprendre le niveau d'intervention de ces processus.

2.1.1. Conditions d'émergence de l'effet phonologique

Lorsque l'on compare les données obtenues pour des groupes tautosyllabiques et hétérosyllabiques de même fréquence, une alternance aléatoire des stimuli de chacune des conditions ne semble pas permettre de faire émerger cet effet. Malgré l'accroissement du nombre de stimuli de remplissage effectué dans l'expérience 5, lequel était censé limiter l'influence des réponses précédentes sur la réponse en cours, aucun effet même minime n'est observé. Or si l'on regroupe les stimuli de chaque condition en début et en fin de liste, l'effet d'alignement attendu se manifeste. Ce changement des résultats lié à un changement de procédure peut s'expliquer par une stabilisation de la structure syllabique ou phonotactique des stimuli-test ; lorsque les stimuli sont mélangés aléatoirement, l'occurrence d'un stimulus CV#CCVC peut induire les participants à attendre un stimulus dont la structure est identique. Si le stimulus suivant correspond au contraire à une structure CVC#CVC, la discordance entre les attentes des participants et la structure effective du stimulus peut conduire à un accroissement de la latence de détection ou à une réponse erronée alors que cette structure est supposée donner lieu à des latences plus courtes et à une quantité moins élevée d'erreurs. Dans la situation inverse, une séquence CVC#CVC suivie d'un stimulus CV#CCVC devrait contribuer à accentuer l'effet attendu. Cependant, la combinaison de ces deux formes d'alternance peut masquer l'émergence de l'effet. Lorsque les conditions sont regroupées en début et en fin de liste, ces phénomènes d'attente sont neutralisés par la régularité de structure des stimuli-test présentés. Il est important de noter que cette procédure aurait tout aussi bien pu conduire à une suppression de l'effet puisque cette situation aurait pu permettre aux participants de prédire la structure des stimuli dans chaque demi-liste. Au contraire, le choix de cette procédure contribue à l'émergence de l'effet.[44]

2.1.2. Caractère écologique des processus impliqués

Il est intéressant de mettre en évidence cet effet phonologique en l'absence de toute contribution possible de la fréquence. La contrainte de procédure nécessaire à l'émergence de cet effet pose cependant le problème du niveau d'intervention de ces connaissances phonologiques. En effet, si les contraintes syllabiques / phonotactiques sont utilisées dans les situations naturelles de traitement de la parole, l'on est en droit de s'attendre à ce qu'un changement de la procédure expérimentale ne modifie pas à ce point les effets observés. Bien sûr, les tâches expérimentales utilisées en psychologie cognitive semblent relativement éloignées des situations naturelles. On demande 'rarement' à un interlocuteur d'appuyer le plus rapidement possible sur un bouton réponse lorsqu'il détecte un mot dans notre énoncé ; lequel est rarement constitué de non-mots qui commencent par des mots. Ces contraintes de la situation expérimentale peuvent cependant être conçues, dans le cas de la tâche de *word-spotting* comme dans celui des

[44] Il convient de noter l'absence d'une telle contrainte de procédure pour la mise en évidence de l'effet d'alignement lorsque la fréquence des groupes de consonnes n'est pas contrôlée. Dans l'expérience 1 en effet, la procédure de mélange aléatoire des stimuli n'empêche pas l'effet de survenir. Il y a certainement ici un lien à établir entre fréquence et structure phonologique. La comparaison de groupes de consonnes différant à la fois en termes de structure syllabique et de fréquence pourrait contribuer plus efficacement à l'émergence de l'effet d'alignement. Il est probable que ces deux catégories d'informations sont utilisées par le système de traitement de la parole. Par ailleurs, les travaux réalisés chez le nouveau-né mettent aussi bien en évidence des effets phonologiques (Friederici & Wessels, 1993 ; Jusczyk, Luce & Charles-Luce, 1994) que des effets de fréquence (Jusczyk et al., 1993). Si ces deux paramètres peuvent correspondre dans certaines expériences à un même phénomène, il est intéressant de poser la question du rôle de la fréquence dans l'acquisition de connaissances qui pourraient donner lieu à l'effet phonologique observé ici. Il est probable que la différence de fréquence des groupes tautosyllabiques et hétérosyllabiques contribue au développement de cette distinction phonologique. Par la suite, ces deux informations pourraient être utilisées –indépendamment ou de manière combinée- dans les processus de traitement de la parole.

paradigmes classiques, comme l'adjonction de processus supplémentaires plus que comme un bouleversement total des procédures de traitement nécessaires. Les énoncés que nous traitons peuvent être considérés comme des séquences sans signification (des non-mots donc) à l'intérieur desquelles les interlocuteurs doivent reconnaître une suite de mots. Ces énoncés ne constituent une séquence de mots qu'une fois les frontières lexicales localisées. L'appui sur un bouton réponse n'est qu'un prolongement au processus d'identification d'un mot dans le signal de parole. Malgré ce parallèle entre le *word-spotting* et les situations naturelles de traitement, la tâche utilisée reste particulièrement difficile. On observe des latences de détection et des taux d'erreur bien plus importants que dans une tâche de détection de phonème. Cette difficulté est probablement liée à l'alternance de non-mots constitués de mots et de non-mots n'en contenant aucun. C'est peut-être l'impossibilité d'identifier des mots réels dans les stimuli de remplissage comme dans la portion finale des stimuli-test qui donne l'impression aux participants que cette tâche est particulièrement artificielle. Du fait de cette difficulté, il est possible que certaines procédures de traitement soient mises en œuvre bien qu'elles ne sont jamais utilisées pour traiter un signal de parole naturel. Alors que nous expliquions les objectifs de l'une des expériences à une étudiante qui venait d'y participer, celle-ci nous a expliqué qu'elle avait cherché à n'extraire que la première syllabe du mot pour accomplir la tâche. Si la première syllabe correspondait à un mot, elle appuyait sur le bouton réponse. Cette procédure est idéale pour faire émerger l'effet, au moins dans l'analyse des taux d'erreur. L'aurait-elle utilisée hors de la situation expérimentale ?

Si la difficulté de la tâche incite les participants à utiliser ce découpage syllabique / phonotactique, rien ne permet d'affirmer que cette procédure est effectivement mise en œuvre dans les situations naturelles de traitement de la parole. Si le découpage phonologique effectué par les participants dans l'expérience 6 n'est que le fruit d'une stratégie de traitement destinée à réussir une tâche difficile, le recours à des connaissances sur la structure phonotactique ou syllabique des séquences de parole ne peut être considéré que comme un processus stratégique post-perceptif qui n'intervient qu'après les processus de traitement de l'information. Si au contraire ce processus est caractéristique des traitements naturels, c'est à dire s'il est effectivement mis en place dans le cadre du traitement de l'information acoustique, phonétique et lexicale, il devrait pouvoir être mis en évidence dans une tâche plus facile que le *word-spotting*. Nous avons donc décidé de reproduire les expériences 3 et 4 en changeant de paradigme expérimental. Vroomen & De Gelder (1999) avaient observé cet effet d'alignement à l'aide d'une tâche de détection de phonèmes. Or cette tâche donne lieu à des latences de détection beaucoup plus courtes et à des taux d'erreur plus faibles que la tâche de word-spotting. C'est donc tout naturellement que nous avons choisi la détection de phonème pour étudier à nouveau la contribution respective de la fréquence et de la structure phonologique dans les processus de segmentation lexicale. Deux expériences ont été conduites avec les mêmes participants. Ceux-ci n'avaient pas pris part aux précédentes. La première est une réplique de l'expérience 3 ; des groupes tautosyllabiques et hétérosyllabiques de même fréquence sont comparés. La seconde reprend les objectifs de l'expérience 4 avec une comparaison de groupes fréquents et rares étant tous hétérosyllabiques.

2.1.3. Expérience 7 : Groupes de consonnes de même fréquence

Dans cette expérience, tout comme dans l'expérience 3, la contribution de la structure phonologique des groupes de consonnes est étudiée indépendamment de la fréquence d'occurrence. Si les processus de segmentation phonotactique ou syllabique sont mis en place dans le cadre des étapes prélexicales du traitement et non comme stratégies de segmentation post-lexicales, l'effet d'alignement observé dans l'expérience 6 devrait émerger à nouveau. Par contre, une absence d'effet dans la tâche de détection de phonèmes conduirait à remettre en cause l'idée d'une intervention prélexicale des processus de segmentation phonologique.

2.1.3.1. Méthode

2.1.3.1.1. Sujets

Vingt-six étudiants en second cycle de psychologie à l'Université Paris 5 - René Descartes ont participé à cette expérience soit volontairement, soit en échange de crédits d'enseignement. Ils étaient tous de langue maternelle française et n'avaient jamais souffert de troubles auditifs. Aucun d'entre eux n'avait pris part aux précédentes expériences. Les mêmes étudiants ont participé à cette expérience et à l'expérience 8.

2.1.3.1.2. Matériel

Le matériel linguistique change notablement par rapport aux stimuli utilisés dans la tâche de word-spotting. Ce sont des groupes nominaux constitués d'un déterminant (l'adjectif indéfini 'un' ou 'une'), d'un nom commun monosyllabique et d'un adjectif qualificatif constitué de deux ou trois syllabes (par exemple 'un sac lavable'). Nous inspirant du travail de Vroomen & De Gelder (1999), nous avons choisi de conduire cette expérience avec des suites de deux mots. Il eut certainement été possible de réutiliser le matériel de word-spotting en changeant la tâche des participants. Il était cependant intéressant d'utiliser un matériel se rapprochant des stimuli couramment rencontrés dans les énoncés de parole naturels. Puisque la problématique que nous souhaitons aborder ici est celle de la distinction entre processus de traitement et stratégies post-lexicales, ceci permet d'accroître la proximité avec une situation naturelle. Ce choix apporte cependant des contraintes nouvelles pour la constitution du matériel. En effet, il convient de sélectionner des suites nom commun + adjectif qui ne soient ni incongrues ni trop prédictibles. Huit mots ont chacun été appariés avec deux adjectifs post-posés. La consonne finale du nom commun est toujours l'occlusive /k/. Associée à l'occlusive finale du nom commun, la consonne initiale de l'adjectif qualificatif constitue un groupe tautosyllabique (lorsque la consonne initiale de l'adjectif est une liquide, par exemple /kl/ dans 'un sac lavable', /œ̃saklavablə/) ou hétérosyllabique (pour les consonnes fricatives ou occlusives, par exemple /kt/ dans 'un sac tissé', /œ̃saktise/). Ces groupes sont appariés sur la base de leur probabilité d'occurrence brute. Aucun des stimuli-test ne contient une voyelle qui permettrait d'influencer le rattachement syllabique de la consonne suivante (comme /o/ ou /e/). Quatre groupes de consonnes différents ont été comparés (cf. Tableau 17).

Tableau 17 : Conditions expérimentales de l'Expérience 7 et détail des groupes de consonnes utilisés.

	Tautosyllabique	**Hétérosyllabique**
Exemple	/œ̃saklavablə/ 'un sac lavable'	/œ̃saktise/ 'un sac tissé'
Types de groupes	occl.-liq.	occl.-fric. occl.-occl.
Groupes consonantiques	/kl/, /kr/	/kt/, /ks/
Probabilités d'occurrence brute (moyenne)	80, 85 (82.8)	82, 85 (83.7)
Probabilités d'occurrence positionnelle (moyenne)	70, 80 (75.6)	1, 21 (12.1)

Tout comme dans les autres expériences, aucune des séquences monosyllabe + consonne contextuelle ne constitue un début de mot possible dans la base de données BRULEX (Content et al., 1990). La même procédure de contrôle de ce critère a été appliquée. Huit séquences de remplissage ont été constituées qui ne contiennent pas la consonne cible /k/ (par exemple 'une robe longue', /ynrɔblɔ̃g/). Huit autres séquences contenaient le phonème-cible dans une autre position (le phonème initial du mot monosyllabique, par exemple 'un comte ruiné', /œ̃kɔ̃trɥine/). L'ordre de passation des expériences 7 et 8 est contrebalancé. La moitié des participants a passé l'expérience 7 en premier, l'autre moitié en premier. La liste des stimuli expérimentaux et de remplissage est présentée en Annexe 18 (p.XXIX).

Les stimuli ont été enregistrés, digitalisés (16 bits, échantillonnage à 16kHz) et stockés sur le disque dur d'un ordinateur de la même manière que dans les précédentes expériences. Un exemplaire de chaque séquence a été sélectionné parmi trois répétitions. Les stimuli ont ensuite été édités afin de mesurer la durée entre le début du stimulus et celui de l'occlusion de la consonne finale du mot monosyllabique. Cette valeur est utilisée afin de mesurer la latence de la réponse à partir de l'occlusion de la consonne cible.

2.1.3.1.3. Procédure

Les participants avaient pour consigne d'appuyer le plus rapidement possible sur un bouton réponse dès qu'ils identifiaient le phonème-cible. Un seul phonème devait être détecté dans cette expérience (/k/). Il pouvait apparaître dans n'importe quelle position mais seules les réponses fournies pour le phonème final du mot monosyllabique sont étudiées. Les temps de réaction étaient mesurés depuis le début du stimulus et corrigés automatiquement par le logiciel de passation afin de refléter la latence entre l'occlusion de la consonne cible et la réponse du sujet. Avant la phase expérimentale proprement dite, les participants se familiarisaient avec la tâche à l'aide d'un matériel différent.

2.1.3.2. Résultats

Du fait du changement de tâche et de son caractère plus facile, des critères de sélection différents ont été utilisés pour choisir les items et les participants qui seraient intégrés dans l'analyse. Le seuil de sélection des participants aussi bien que des items ont été fixés *a priori* à 70 % de détections correctes (au lieu de 50 %). Tout comme avec le paradigme de *word-spotting*, les latences de détection correspondant à la moyenne ± deux écart-types ne sont pas intégrées dans les analyses. Les données de deux sujets n'ont pu être prises en compte en raison de latences de détection ou d'omissions trop importantes. Tous les items présentés sont intégrés. Les données non-intégrées dans l'analyse des latences de détection constituent 5.2 % de l'ensemble des observations.

2.1.3.2.1. Temps de réaction

Les temps de réaction et les taux d'erreur observés sont présentés dans le Tableau 18. Les variables *Liste* et *Ordre* n'interagissent pas avec la variable *Contexte*. L'interaction de second ordre n'est pas significative. Les analyses sont conduites avec le *Contexte* (C_2) comme variable intra-sujet ($S * C_2$) et intra-item ($I * C_2$). La latence moyenne de détection est de 601 ms (SD = 183 ms). Les latences obtenues pour les groupes tautosyllabiques et hétérosyllabiques sont tout à fait similaires (respectivement 604 ms et 598 ms). Cette différence n'est pas significative

($F_1 < 1$; $F_2 < 1$). Contrairement aux données observées dans la tâche de word-spotting de l'expérience 6, l'analyse des latences de détection ne permet de mettre en évidence aucun effet du contexte phonologique. On retrouve ici une absence d'effet similaire à ce qui avait été observé dans les expériences 4 et 5.

Tableau 18 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 7 en fonction du type de groupe consonantique.

	Tautosyllabique	**Hétérosyllabique**
TR	604 (41)	598 (34)
Taux d'erreur	3.1	7.3

2.1.3.2.2. Taux d'erreur

Les mêmes analyses ont été conduites sur les taux d'erreur (5.2 % de l'ensemble des données). Les variables *Liste* et *Ordre* n'interagissent pas avec la variable *Contexte*. L'interaction de second ordre n'est pas non plus significative. Les groupes tautosyllabiques donnent lieu à des taux d'erreur légèrement *moins* élevés (3.1 %) que les groupes hétérosyllabiques (7.3 %). Le sens de cette différence est contraire à ce qui peut être attendu si l'hypothèse d'un recours à des connaissances sur la structure phonologique des groupes de consonnes dans les processus de segmentation lexicale est valide. Cette différence n'est cependant pas significative ($F_1(1,23) = 1.643$, $p > .1$; $F_1(1,7) = 2.333$, $p > .1$).

2.1.3.3. Discussion

Les données de l'expérience 7 reproduisent l'absence d'effet déjà observée dans les expériences 3 et 5. Contrairement à ce qui avait été observé dans l'expérience 6, le recours à la tâche de détection de phonèmes plutôt qu'à celle de *word-spotting* ne permet pas de mettre en évidence un effet de l'alignement entre frontières phonologique et lexicale. Ce changement de tâche a été effectué afin de répondre à la question essentielle du niveau d'intervention des connaissances sur les contraintes phonologiques de la langue dans les processus de segmentation de la parole en mots. Nous avions formulé l'hypothèse que si ces représentations phonologiques sont utilisées au cours des étapes précoces du traitement, l'effet devrait émerger dans n'importe quelle situation expérimentale *ad hoc*. Au contraire, si l'effet observé dans l'expérience 6 peut relever de processus décisionnels, l'adoption d'une tâche beaucoup plus facile aurait du donner lieu à une disparition de l'effet. La comparaison des taux d'erreurs mesurés dans les précédentes expériences de *word-spotting* et dans la tâche de détection de phonèmes présentée confirme la différence de difficulté des deux tâches. En *word-spotting*, on observe des taux d'erreurs qui oscillent entre 15.0% (Expérience 1) et 23.9 % (Expérience 5). La proportion d'erreurs observée dans l'expérience 6 est nettement plus faible (5.2 %). L'adoption d'une tâche moins difficile que le *word-spotting* semble donc contribuer à l'abandon de procédures de segmentation phonologique de la chaîne de parole ; ce qui semble confirmer leur caractère artificiel. Il convient cependant, avant de conclure à une implémentation post-perceptive de l'effet phonologique observé dans l'expérience 6, d'acquérir la certitude que des effets mis en évidence avec le paradigme de *word-spotting* peuvent être répliqués dans une tâche de détection de phonèmes. Dans l'expérience 4, la comparaison de groupes de consonnes hétérosyllabiques fréquents et rares donne lieu à des taux d'erreurs qui diffèrent significativement dans l'analyse par sujet. Cet effet se manifeste malgré une procédure tout à fait similaire à celle de l'expérience 3 dans laquelle aucun effet n'émerge ; il est donc judicieux de répliquer l'expérience 4 avec un matériel similaire.

2.1.4. Expérience 8 : Estimation du rôle de la fréquence

Cette expérience est une réplique de l'expérience 4 avec un matériel linguistique différent adapté à la tâche de détection de phonèmes. Elle a pour objet d'évaluer la contribution de la fréquence des groupes de consonnes dans les processus de segmentation lexicale et de déterminer dans quelle mesure la tâche de détection de phonème permet de faire émerger l'effet observé dans l'analyse par sujet des données de l'expérience 4. Comme dans l'expérience originale, les groupes comparés sont tous hétérosyllabiques.

2.1.4.1. Méthode

2.1.4.1.1. Sujets

Les participants sont ceux qui ont pris part à l'expérience 7.

2.1.4.1.2. Matériel

Les caractéristiques du matériel linguistique sont similaires à celles des stimuli de la précédente expérience. Ce sont des groupes nominaux constitués d'un déterminant (l'adjectif indéfini 'un' ou 'une'), d'un nom commun monosyllabique et d'un adjectif qualificatif constitué de deux ou trois syllabes. Douze mots ont chacun été appariés avec deux adjectifs post-posés. La consonne finale du nom commun est toujours l'occlusive /p/ ; c'est la consonne-cible. Associée à l'occlusive finale du nom commun, la consonne initiale de l'adjectif qualificatif génère un groupe fréquent (par exemple /ps/ dans 'une coupe soldée', /ynkupsɔlde/) ou rare (par exemple /pʃ/ dans 'une coupe chinoise', /ynkupʃinwaz/). Ces groupes sont tous hétérosyllabiques. Aucun des stimuli-test ne contient une voyelle qui permettrait d'influencer le rattachement syllabique de la consonne suivante (comme /o/ ou /e/). Quatre groupes de consonnes différents ont été comparés (cf. Tableau 19).

Tableau 19 : Conditions expérimentales de l'Expérience 8 et détail des groupes de consonnes utilisés.

	Fréquent	**Rare**
Exemple	/ynkupsɔlde/ 'une coupe soldée'	/ynkupʃinwaz/ 'une coupe chinoise'
Types de groupes	occl.-fric. occl.-occl.	occl.-fric. occl.-occl.
Groupes consonantiques	/ps/, /pt/	/pʃ/, /pk/
Probabilités d'occurrence brute (moyenne)	75, 72 (73.5)	6, 21 (13.5)
Probabilités d'occurrence positionnelle (moyenne)	57, 1 (29)	1, 1 (1)

Tout comme dans les autres expériences, aucune des séquences monosyllabe + consonne contextuelle ne constitue un début de mot possible dans la base de données BRULEX (Content et al., 1990). Douze séquences de remplissage ont été constituées qui ne contiennet pas la consonne cible /p/ (par exemple 'une robe longue'). Douze autres séquences contenaient le phonème-cible dans une autre position (le phonème initial du mot monosyllabique, par exemple 'un comte ruiné'). La liste des stimuli expérimentaux et de remplissage est présentée en Annexe 19 (p.XXX). L'ordre de passation des expériences 7 et 8 était contrebalancé entre participants.

Les stimuli ont été enregistrés, digitalisés (16 bits, échantillonnage à 16kHz) et stockés sur le disque dur d'un ordinateur de la même manière que dans les précédentes expériences. Un exemplaire de chaque séquence a été sélectionné parmi trois répétitions. Les stimuli ont ensuite été édités afin de mesurer la durée entre le début du stimulus et celui de l'occlusion de la consonne finale du mot monosyllabique. Cette valeur est utilisée afin de mesurer la latence de la réponse à partir de l'occlusion de la consonne-cible portée par le monosyllabe.

2.1.4.1.3. Procédure

Les participants avaient pour consigne d'appuyer le plus rapidement possible sur un bouton réponse dès qu'ils identifiaient le phonème-cible. Un seul phonème devait être détecté dans cette expérience (/p/). Il pouvait apparaître dans n'importe quelle position mais seules les réponses fournies pour le phonème final du mot monosyllabique sont étudiées. Les temps de réaction étaient mesurés depuis le début du stimulus et corrigés automatiquement par le logiciel de passation afin de refléter la latence entre l'occlusion de la consonne cible et la réponse du sujet. Avant la phase expérimentale proprement dite, les participants se familiarisaient avec la tâche à l'aide d'un matériel différent.

2.1.4.2. Résultats

Les critères de sélection des participants et des items sont identiques à ceux de l'expérience 7. Les données des deux participants qui n'avaient pas atteint le critère dans la précédente expérience ne sont pas prises en compte. A l'exception des ces deux participants, aucun autre ne présente des taux d'omissions supérieurs au critère de 30 % dans cette expérience. Tous les items présentés sont intégrés dans les analyses. Les temps de réaction supérieurs à 944 ms n'ont pas été pris en compte.. Les données non-intégrées dans l'analyse des latences de détection constituent 5.2 % de l'ensemble des observations.

2.1.4.2.1. Temps de réaction

Les temps de réaction et les taux d'erreur observés sont présentés dans le Tableau 20. La variable *Liste* n'interagit pas avec la variable *Contexte* ; elle n'est donc pas intégrée dans les analyses de variance qui ont été effectuées sur les temps de réaction. L'ordre de passation des expériences 7 et 8 n'interagit pas non plus avec les autres variables. L'effet d'interaction de second ordre n'est pas significatif. Les analyses de variance sont conduites avec le *Contexte* (C_2) comme variable intra-sujet ($S * C_2$) et intra-item ($I * C_2$). La latence moyenne de détection est de 483 ms (SD = 115 ms). Les latences de détection obtenues pour les groupes fréquents sont plus élevées (501 ms) que pour les groupes rares (465 ms).

Tableau 20 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 8 en fonction de la fréquence du groupe consonantique.

	Fréquent	**Rare**
TR	501 (25)	465 (21)
Taux d'erreur	4.3	6.1

Cette différence est significative dans l'analyse par sujet ($F_1(1,23) = 6.371$, $p < .05$) ainsi que dans celle par item ($F_2(1,11) = 10.716$, $p < .01$). La détection du phonème final d'un mot monosyllabique est donc plus lente lorsque celui-ci est prononcé au début d'un groupe fréquent que lorsqu'il forme, avec le phonème contextuel, un groupe relativement rare dans la langue. Cet effet est tout à fait similaire à celui qui avait été observé dans une tâche de word-spotting -uniquement dans l'analyse par sujet- avec un matériel semblable (expérience 4).

2.1.4.2.2. Taux d'erreur

Les mêmes analyses ont été conduites sur les taux d'erreur (5.2 % de l'ensemble des données). Les variables *Liste* et *Ordre* n'interagissent pas avec la variable *Contexte*. L'interaction de second ordre n'est pas non plus significative. La différence de taux d'erreur entre groupes fréquents (4.3 %) et rares (6.1 %) n'atteint pas le seuil de significativité ($F_1 < 1$; $F_2 < 1$).

2.1.5. Discussion générale

Les tâches de *word-spotting* et de détection de phonèmes fournissent des informations assez contradictoires. Alors que les données de la dernière expérience de *word-spotting* contribuaient à attester -conformément aux résultats obtenus par McQueen (1998) et par Vroomen & De Gelder (1999)- le rôle des régularités phonologiques dans les processus de segmentation lexicale, les données de la tâche de détection de phonème font uniquement émerger un effet de la fréquence des groupes de consonnes et infirment en ce sens l'hypothèse d'un recours aux connaissances sur les régularités phonotactiques de la langue dans les processus de segmentation de la parole. Ce changement de tâche a été décidé afin de vérifier que l'effet observé dans l'expérience 6 pouvait être répliqué dans une situation plus facile et -peut-être- plus naturelle que le paradigme de word-spotting. L'hypothèse d'une implémentation post-lexicale de l'effet semble de voir être privilégiée. En effet, si l'effet d'alignement entre frontières phonologique et lexicale n'émerge que dans une tâche relativement difficile, disparaissant totalement lorsque la difficulté de la tâche est réduite, il est peu probable que ces procédures de découpage du signal fondées sur les régularités phonologiques de la langue soient partie intégrante des processus précoces de traitement de la chaîne parlée. Il semble plus approprié de concevoir ces processus comme un support auquel auraient recours les auditeurs dans le cadre de la situation expérimentale ; ce qui conduirait à limiter considérablement leur apport potentiel dans les processus de reconnaissance de la parole. Deux points doivent cependant être abordés ici afin de ne pas conclure trop rapidement au caractère artefactuel des effets phonologiques observés dans les études précédentes. Le premier concerne le statut de la tâche qui, par définition, ne suppose d'avoir accès qu'à des représentations segmentales : les phonèmes. Le second repose sur l'inversion de l'effet observée entre les expériences 1 et 2 ; cette observation soulève également la question de la pertinence de la tâche de détection de phonèmes dans l'étude de processus de segmentation fondés sur des représentations phonologiques plus grandes que le phonème.

Le cœur du travail présenté ici consiste à poser la question du rôle éventuel d'informations dont le support acoustique excède le phonème dans les processus de perception de la parole. Les régularités phonologiques étudiées sont liées à l'agencement des phonèmes entre eux. Que l'on parle de régularités phonotactiques ou de structure syllabique, ce sont les caractéristiques des *séquences* de phonèmes qui déterminent leur statut. Une suite de phonèmes constitue une *bonne forme* en début de mot, est prononcée en *attaque* syllabique, en fonction d'une certaine relation entre les segments phonémiques dont elle est composée. Dans les études présentées, les caractéristiques d'un segment ne déterminent pas son rattachement à l'arborescence syllabique ou son degré d'acceptabilité en début de mot[45]. Les seules représentations valides pour l'implémentation de procédures de segmentation de la parole fondées sur ce type d'informations reposent sur les régularités de *co-occurrence* des segments qui les constituent. Or la tâche de détection de phonèmes peut tout à fait être réalisée sans faire intervenir ce type de connaissances contextuelles. La tâche des participants consiste essentiellement, à partir d'une cible prédéterminée, à identifier un phonème -donc un *segment*- dans le signal acoustique présenté. Dans la tâche de *word-spotting* au contraire, l'information à extraire est constituée de plusieurs phonèmes ; l'objectif primordial étant de dissocier deux ensembles de phonèmes de manière à ce que l'un d'entre eux constitue un mot de la langue. Alors que la tâche de détection de phonèmes implique l'accès à des représentations phonémiques, celle de word-spotting ne nécessite pas d'identifier -en tout cas consciemment- les phonèmes qui constituent ce signal de parole. Plus précisément, il n'est pas certain que le type de représentation phonémique développé dans la tâche de détection de phonèmes soit identique à celui qui se met en place lorsque l'on traite un signal de parole afin d'identifier les mots dont il est constitué (cf. Foss & Blank, 1980 ; Segui et al., 1981 ; Szende, 1980 pour des discussions plus approfondies sur ce thème). L'on est alors en mesure d'envisager que les informations sur lesquelles repose la décision de détection du phonème-cible ne sont pas nécessairement influencées par des représentations suprasegmentales. La tâche de détection de phonème pourrait alors être considérée comme *moins écologique* que celle de *word-spotting* ; ce qui expliquerait l'absence d'effet d'alignement entre frontières suprasegmentale et lexicale. Les données présentées par Vroomen & De Gelder (1999) avec une

[45] Il convient cependant de mentionner qu'à l'inverse, les caractéristiques acoustiques d'un phonème sont grandement déterminées par sa position dans la chaîne de parole (cf. par exemple Autesserre & Chafcouloff, 1999 pour une étude des caractéristiques du /r/ en fonction de sa position dans la syllabe). Malgré la discussion menée dans le cadre de l'expérience 2, ce problème n'a pas été approfondi dans le cadre de cette thèse. Il est cependant ➥

tâche de détection de phonèmes vont pourtant dans le sens d'un rôle des processus de segmentation phonologique dans le traitement de la parole. La discussion développée dans les chapitres 3 et 4 montre toutefois que l'effet observé peut s'expliquer par des paramètres probabilistes plutôt que phonologiques. En effet, s'il est avéré que la comparaison de groupes de consonnes se distinguant sur des paramètres à la fois phonologiques et probabilistes peut donner lieu à un effet en détection de phonèmes, rien ne permet d'affirmer qu'une différence d'ordre purement phonologique permet d'observer un effet dans cette tâche. Rappelons que l'émergence d'un effet phonologique dans la tâche de détection de phonèmes repose sur une conception lexicale du processus de détection phonémique (cf. section 2.2.3.2 du Chapitre 2). Si le phonème-cible est plus long à détecter lorsqu'il est à l'attaque de la syllabe suivante, c'est parce que le mot auquel il appartient serait plus long à identifier. Le délai introduit dans l'identification du mot introduirait un délai dans l'identification du phonème-cible. Puisque dans le travail de Vroomen & De Gelder (1999), la fréquence n'est pas contrôlée, l'effet qui est observé dans la tâche de détection de phonème mise en œuvre par les auteurs pourrait très bien avoir comme source unique les phénomènes de compétitions lexicales ; c'est également le cas pour les données de l'expérience 8 qui pourraient tout à fait refléter un effet du déséquilibre dans le nombre de candidats lexicaux activés par les groupes fréquents et rares. Si les informations de structuration phonologique de la chaîne parlée ne sont pas utilisées dans la tâche de détection de phonèmes, c'est peut-être simplement parce que ce paradigme fait intervenir des processus d'identification phonémique qui ne peuvent pas reposer sur des informations suprasegmentales.

La discussion des résultats de l'expérience 2 était centrée sur le rôle possible de différences acoustiques entre le phonème initial des groupes tautosyllabiques et celui des groupes hétérosyllabiques. Dans cette expérience de décision lexicale, l'effet observé en *word-spotting* s'inverse. Alors que les latences de détection mesurées dans l'expérience 1 étaient plus longues pour les groupes occlusive-liquide que pour les deux autres catégories, on observe des taux d'erreur qui sont quant à eux plus faibles pour les groupes occlusive-liquide que pour les groupes occlusive-fricative et occlusive-occlusive. Nous avons émis l'idée que cette 'inversion' pouvait s'expliquer par un effet articulatoire du contexte qui conduirait à une réduction acoustique plus importante de l'occlusive lorsqu'elle est prononcée devant une fricative ou une occlusive que devant une liquide. Cette réduction acoustique influencerait les données de la tâche de décision lexicale mais ne serait pas suffisante pour masquer l'effet contextuel dans la tâche de *word-spotting*. Si l'on rapporte cette proposition aux données de la tâche de détection de phonème, il

---

évident qu'il constitue un point essentiel de la compréhension des processus de reconnaissance de la parole.

est possible que les processus impliqués soient plus sensibles à ce phénomène de modification acoustique en fonction du contexte que ceux qui sont mis en œuvre en *word-spotting*. L'effet acoustique pourrait alors masquer un autre effet : celui de l'alignement entre frontières suprasegmentale et lexicale. Cette compensation des deux effets mettrait en échec l'étude des procédures de segmentation phonotactique / syllabique que l'on observe dans l'expérience 1.

Il est par conséquent essentiel de ne pas juger trop vite de l'absence d'effet phonologique dans la tâche de détection de phonèmes. Les deux problèmes soulevés nous semblent cependant contradictoires. Si comme nous l'avons supposé dans la première partie de cette discussion, la détection du phonème-cible ne fait pas intervenir les processus de segmentation suprasegmentale, pourquoi l'effet de réduction acoustique ne se manifeste-t-il pas ? Parallèlement, il a été montré que les représentations phonémiques développées dans la tâche de détection de phonèmes sont elles-mêmes dérivées des représentations lexicales. Il est alors légitime de se demander pourquoi l'accès à ces représentations lexicales n'est pas influencé par la structure phonotactique ou syllabique de la chaîne de phonèmes. Peut-être parce que ces procédures de segmentation ne font pas partie intégrante des traitements effectués sur le signal de parole mais n'interviennent qu'après les étapes perceptives… Des travaux complémentaires sont nécessaires afin d'approfondir cette problématique.

2.2. Interaction entre variables prélexicales et lexicales

Le second problème posé par les données obtenues dans ce travail se rapporte à la restriction des effets phonologiques observés. Chaque fois qu'un effet d'alignement entre frontières phonologique et lexicale a été mis en évidence, il s'est restreint à un sous ensemble du matériel. Cette restriction soulève à nouveau, bien que sous un angle différent, la question du niveau d'implémentation des processus impliqués dans les effets observés. Si la prise en compte d'informations prélexicales fondées sur les caractéristiques des groupes de consonnes médians n'est possible qu'en fonction de certains paramètres lexicaux, cette information censée être utilisée au cours d'étapes précoces du traitement est peut-être mise en œuvre après les étapes perceptives et / ou d'accès au lexique. En effet, les informations prélexicales sont supposées intervenir avant même la mise en place des processus lexicaux. Pour que l'effet d'une variable prélexicale puisse dépendre de paramètres lexicaux, il faudrait que cette variable soit en réalité prise en compte à une étape ultérieure. Il est cependant possible que ce qui a été considéré comme une régularité lexicale corresponde en réalité à une régularité sublexicale, laquelle pourrait alors entrer en interaction avec les processus prélexicaux de segmentation phonologique.

Cette possibilité découle du lien étroit qui peut être mis en évidence entre le nombre de voisins lexicaux d'un mot et la fréquence des diphones qui le constituent.

Deux cadres explicatifs peuvent être proposés pour rendre compte de la restriction de l'effet d'alignement à un sous-ensemble des mots de l'échantillon. La dichotomie effectuée sur base lexicale (nombre de compétiteurs ou fréquence des mots) dans les diverses analyses complémentaires présentées pourrait correspondre à des spécificités du matériel qui relèvent en réalité de caractéristiques sublexicales. L'opposition entre mots ayant beaucoup ou peu de compétiteurs pourrait ainsi correspondre à une autre dichotomie qui, quant à elle, reposerait sur des régularités phonémiques ou phonologiques des mots utilisés. Cette explication relève d'une éventuelle confusion entre des caractéristiques lexicales et sublexicales. En outre, tout paradigme expérimental nécessite d'obtenir une réponse de la part des participants. Ceux-ci doivent décider, à partir des informations qui leur sont disponibles, quelle réponse ils doivent donner. Les participants doivent donc utiliser toutes les informations dont ils disposent pour prendre une décision. Il est possible que plusieurs catégories d'informations soient comparées au cours de l'étape de prise de décision afin de donner la meilleure réponse possible, même si ces informations correspondent à des niveaux de traitement différents.

2.2.1. Nombre de voisins lexicaux et fréquence des diphones

Vitevitch, Luce, Pisoni, & Auer (1999) montrent que les mots qui ont une densité de voisinage lexical élevée[46] sont souvent constitués de suites de phonèmes fréquentes dans la langue. La variable que nous avons utilisée pour dichotomiser le matériel de l'expérience 1 n'est pas le 'nombre de voisins lexicaux' tel qu'il est décrit par Vitevitch, Luce, Pisoni, & Auer (1999). Lorsque nous parlons ici de 'nombre de compétiteur', c'est à partir d'un calcul prenant en compte le nombre de mots *commençant* par la séquence de phonèmes d'un mot donné. Le 'nombre de voisins lexicaux' correspond au contraire à la quantité de mots du lexique qui partagent un ou plusieurs phonèmes, et ce quelle que soit leur position dans les mots. Ces deux méthodes de calcul sont tout aussi pertinentes l'une que l'autre. Dans le cas d'un modèle comme COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987), c'est le 'nombre de compétiteurs' tel que nous l'avons calculé qui influence les processus de reconnaissance. Des modèles comme TRACE (McClelland & Elman, 1986) ou le *Neighborhood Activation Model* (Luce & Pisoni, 1998) font au contraire intervenir des processus qui reposent essentiellement sur le 'nombre de voisins lexicaux' puisque les candidats activés peuvent l'être en tout endroit de la

[46] C'est à dire qu'ils partagent 1 à N phonèmes avec une quantité importante de mots.

chaîne de phonèmes. En réalité, il est probable que ces deux catégories de processus soient utilisées par le système cognitif pour traiter les signaux de parole (Mattys, 1997). Quoiqu'il en soit, il est essentiel de noter qu'un observable envisagé sous un angle lexical (le nombre de mots avec lesquels un mot peut entrer en compétition par exemple) peut tout à fait se confondre avec un autre observable, lequel repose sur des représentations uniquement sublexicales comme la fréquence des diphones. Nous avons déjà abordé cette question dans le cadre du Chapitre 1, ce qui nous avait permis de mettre en évidence la difficulté à dissocier des variables qui, souvent, se confondent dans l'étude des processus psycholinguistiques. Observer une interaction entre des variables respectivement sublexicale et lexicale peut donc s'expliquer en réalité par une interaction faisant intervenir deux variables sublexicales. En ce qui concerne la restriction de l'effet du contexte phonologique à un ensemble de mots partageant des caractéristiques lexicales (le nombre de compétiteurs ou la fréquence), on peut imaginer que ce que nous avions considéré comme une variable lexicale peut désigner une spécificité des mots de notre échantillon qui pourrait en réalité s'exprimer à un niveau de traitement prélexical et interagir avec les processus de segmentation phonologique étudiés.

Si l'influence du nombre de compétiteurs est envisagée sous son angle lexical, il conviendrait de considérer l'effet du contexte phonologique observé dans l'Expérience 1 comme la manifestation d'une effet dont la source serait à rechercher au niveau du lexique (cf. le lien entre fréquence et compétitions lexicales) ou à une étape post-perceptive / décisionnelle. Si au contraire l'influence du nombre de compétiteurs sur l'émergence de l'effet de contexte s'explique par des caractéristiques sublexicales des mots, il serait possible d'envisager cette restriction de l'effet observé dans la première expérience comme une interaction entre deux variables intervenant au cours des étapes prélexicales du traitement. Cette seconde solution constituerait donc un argument en faveur d'une influence purement prélexicale des effets phonologiques observés. Nous avons donc cherché à identifier des régularités structurelles des items de l'Expérience 1 qui correspondraient à la dissociation que nous avons opérée entre les mots qui ont peu ou beaucoup de compétiteurs. Cette régularité existe et se manifeste dans la structure consonne-voyelle des mots utilisés. Dans le matériel que nous avions choisi, tous les mots se terminaient par une séquence VC. Nous avions cependant mélangé des mots *commençant* par une consonne unique (CVC) avec d'autres dont l'initiale formait un groupe de consonnes (CCVC). Les mots qui ont beaucoup de compétiteurs ont presque tous une structure CVC (16 sur 18) alors que c'est l'inverse pour ceux qui ont peu de compétiteurs (4 CVC et 9 CCVC). Rappelons que les suites CC sont nettement plus rares dans la langue que les consonnes intervocaliques (cf. Chapitre 4). Il est donc possible que nous ayons proposé une interprétation lexicale de la restriction de l'effet alors qu'elle est explicable en termes probabilistes donc prélexicaux ; deux variables prélexicales se combineraient alors, l'une reposant sur la fréquence des diphones à l'intérieur des mots et l'autre sur le statut phonologique du groupe de consonnes médian. Dans l'expérience 6, c'est la restriction fondée sur la fréquence d'usage des mots qui contribue à faire émerger un effet stable du type de groupe consonantique. Une analyse du matériel des expériences 3, 5 et 6[47] n'a cependant pas permis de mettre au jour une régularité similaire entre les paramètres lexicaux liés à la fréquence des mots de l'échantillon et des régularités sublexicales comme la fréquence des diphones. La structure Consonne-Voyelle des mots de l'échantillon est équitablement distribuée entre les mots les plus fréquents (6 CVC, 3 CCVC) et les plus rares (5 CVC, 3 CCVC) de l'échantillon. Les données de cette expérience ne permettent pas de confirmer les conclusions avancées par l'étude de la structure phonémique du matériel de l'expérience 1. Cette expérience nous laisse donc dans l'incertitude concernant le lien entre la variable lexicale que constitue la fréquence des mots et la variable sublexicale supposée à l'œuvre dans la segmentation phonologique du signal de parole. Il est possible que nous n'ayons pas su trouver un paramètre sublexical adéquat dans le matériel de l'expérience 6. Il est aussi possible que l'on soit effectivement en présence d'une restriction de l'effet fondée sur des informations lexicales ; ce qui inciterait à concevoir l'effet d'alignement observé dans l'expérience 6 comme un effet post-perceptif influencé par les traitements lexicaux. Cette conclusion serait conforme à l'absence d'effet d'alignement entre frontières phonologique et lexicale dans la tâche de détection de phonèmes. Une autre interprétation se doit cependant d'être mentionnée. Pour cela, il convient d'aborder une réflexion sur l'activité cognitive des auditeurs lorsqu'ils traitent un signal de parole et, plus particulièrement, lorsqu'ils réalisent une tâche expérimentale.

2.2.2. Intégration des informations

Lorsqu'un participant est confronté à une tâche, l'expérimentateur attend de lui qu'il fournisse pour chaque essai une réponse conforme à la consigne qui lui a été donnée. Cette réponse repose sur une série de processus qui doivent être modélisés au mieux afin de pouvoir décrire et interpréter l'influence des variables manipulées. Dans la tâche de word-spotting, les participants sont soumis à un signal acoustique qu'ils doivent découper pour identifier un mot. Ce signal acoustique doit être analysé et recodé en une chaîne de segments phonologiques (séquences de spectres à court terme, matrices de traits, phonèmes, syllabes). Ces segments seront appariés avec les représentations stockées dans le lexique afin d'identifier les mots

[47] Pour lesquelles les stimuli-test sont les mêmes.

possibles. L'hypothèse proposée par McQueen (1998) et par Vroomen & De Gelder (1999) est que ces régularités phonologiques peuvent être utilisées par les locuteurs lorsqu'ils traitent un signal de parole. La mise en œuvre de ces connaissances dans les processus de traitement de la parole pourrait jouer un rôle dans la localisation des frontières de mots. Les données expérimentales ne confirment qu'en partie cette proposition. Plus exactement, elles confirment l'influence de ces processus mais ne permettent pas de fournir une réponse satisfaisante à la question de leur niveau d'implémentation. Il est cependant important de comprendre pourquoi certains mots ne permettent pas de faire émerger cet effet. On peut supposer que le recours à ces régularités sera mis en place avant les processus de traitement lexicaux et que leur déroulement dépendra de ces régularités. C'est ce qui est proposé par Norris, McQueen, Cutler, & Butterfield (1997 ; Suomi, McQueen, & Cutler, 1997). Selon eux, le découpage effectué sur la base des régularités phonologiques de la langue servirait à moduler les niveaux d'activation lexicale des candidats. Selon cette conception, l'effet des régularités phonologiques ne devrait pas dépendre de caractéristiques lexicales, sauf si l'implémentation de ces processus de segmentation phonologique est post-perceptive.

On peut au contraire envisager que ces deux processus sont relativement indépendants. Le décours temporel des activations lexicales pourrait s'effectuer sans qu'aucune modification ne soit introduite par le processus de segmentation phonotactique ou probabiliste. Par contre, ces informations pourraient être utilisées comme une source d'information supplémentaire pour décider de la présence d'un mot dans le stimulus. On sait que certains mots sont plus faciles à reconnaître que d'autres. Diverses caractéristiques aussi bien lexicales que prélexicales peuvent influencer la durée nécessaire ou le rapport Signal / Bruit acceptable pour être en mesure d'identifier un mot. Ces caractéristiques peuvent être mises en évidence dans des tâches comme la décision lexicale ou l'identification de mot dans du bruit, paradigmes dans lesquels le découpage lexical est déjà présent puisque ce sont des mots isolés qui sont présentés. En fonction de la difficulté intrinsèque des mots pour être reconnus comme présents dans le lexique, cette difficulté reposant sur la fréquence d'usage, le nombre de voisins lexicaux, le nombre de candidats maintenus dans la cohorte ou la fréquence des diphones qui les constituent, l'utilisation de procédures de segmentation phonotactique ou syllabique pourrait n'être pas efficace. On peut par exemple supposer que si des mots sont très faciles à reconnaître, la mise en œuvre de connaissances sur les régularités de la langue sera -dans une tâche de *word-spotting*- de peu d'intérêt pour localiser la frontière entre le mot et le reste du non-mot. Les informations utilisées pour identifier le mot en isolation pourraient suffire. Au contraire, des mots intrinsèquement difficiles à identifier pourraient bénéficier d'indices supplémentaires pour que la frontière entre le mot et le reste du non-mot soit localisée. Cette explication pourrait rendre compte de la restriction de l'effet aux mots qui ont beaucoup de compétiteurs dans l'expérience 1. Le modèle NAM (Luce & Pisoni, 1998) prédit que les mots qui ont beaucoup de voisins lexicaux seront plus difficiles à identifier parce qu'ils entrent en compétition avec une quantité plus élevée de candidats. Selon notre proposition, cette difficulté permettrait au système de recourir à des processus de segmentation phonotactique ou syllabique alors que ces mêmes processus n'influenceraient pas les latences de décision ou les taux d'erreur lorsque le mot entre en compétition avec une faible quantité de candidats. On retrouve ici des propositions similaires à celles qui étaient faites dans le cadre du modèle RACE (Cutler & Norris, 1979). Dans ce modèle (cf. 2.1 du Chapitre 1), les informations lexicales n'influencent les processus d'identification phonémique que si l'accès au lexique est 'plus rapide' que l'accès à une représentation phonémique abstraite qui prendrait appui sur une voie non-lexicale[48]. La restriction de l'effet aux mots fréquents (expérience 6) semble plus difficile à expliquer. Les mots fréquents sont en effet plus faciles à identifier que les mots rares. Ceci n'est cependant vrai que toutes choses égales par ailleurs. Or l'on peut tout à fait envisager que la distinction entre mots fréquents et rares dans notre échantillon correspond à une autre distinction[49], par exemple le nombre de 'voisins lexicaux'. Si les mots fréquents ont plus de voisins lexicaux que les mots rares, l'effet du nombre de voisins lexicaux devrait contribuer, au contraire de la fréquence, à rendre plus difficile leur identification. Tout dépend alors de l'interaction entre ces deux variables. Si la fréquence ne suffit pas à compenser l'effet du nombre de voisins lexicaux, les mots fréquents risquent d'être plus difficiles à identifier que les mots rares ; ce qui permettrait de fournir la même explication à la restriction de l'effet dans l'expérience 6.

3. Mode d'intégration

La problématique de l'intégration des informations invite naturellement à conclure ce travail par une discussion sur le mode d'intégration des processus de segmentation phonotactique / syllabique dans le cadre des procédures de traitement de la parole. La question abordée ici est celle de l'influence des processus de segmentation prélexicale sur les processus d'identification des mots. Deux axes théoriques peuvent être décrits. Ils découlent de deux

[48] Cf. la distinction entre les représentations phonologiques dérivées de la reconnaissance des mots dans la tâche de détection de phonèmes et les représentations phonétiques qui sont à l'interface de l'image auditive et des représentations lexicales et dont l'objet est d'en permettre l'appariement.

conceptions du lien entre les étapes prélexicale et lexicale du traitement de la parole. Le premier cadre théorique découle directement du paradigme[50] connexionniste et fait reposer l'émergence de l'effet d'alignement sur le niveau d'activation des unités lexicales. Le second repose sur des propositions et des observations diverses dans le champ de la psychologie cognitive, que ce soit dans le cadre spécifique de la psycholinguistique ou dans les domaines plus éloignés de l'étude des processus perceptifs tels que la vision ou le traitement musical.

3.1. Approches théoriques

3.1.1. Influence directe des processus prélexicaux sur les niveaux d'activation lexicale

Dans le cadre du paradigme connexionniste, les travaux sur la reconnaissance des mots attribuent un statut essentiel à la notion de *niveau d'activation des unités lexicales* (Frauenfelder & Peeters, 1990; McClelland & Elman, 1986; Norris, 1994). Puisque la tâche du système consiste à reconnaître les mots de la langue, la comparaison des résultats de simulation avec les données comportementales repose sur l'évolution dans le temps du niveau d'activation des représentations lexicales. Alors que les premiers modèles reposent sur le principe d'une segmentation lexicale conçue comme conséquence des processus de compétition entre candidats, l'introduction de procédures de segmentation prélexicales repose elle aussi sur la modification du niveau des activations lexicales.

Frauenfelder & Peeters (1990) présentent des simulations du modèle TRACE (McClelland & Elman, 1986) dans lesquelles des indices de frontière prélexicaux sont insérés en position médiane d'un mot plurisyllabique ; chacun des segments générés correspondant à un mot dans le lexique du modèle (par exemple, le modèle reçoit /kæt - lɔg/ en entrée ; les mots *cat*, *log* et *catalog* font partie du lexique de simulation). L'indice de segmentation a le même statut qu'un phonème. Ils observent que l'insertion de ces marques de frontières ne permet pas au modèle de reconnaître les deux mots enchâssés : c'est le mot porteur, alors que l'entrée phonémique comporte une frontière, qui est reconnu. Les processus implémentés dans TRACE ne semblent donc pas permettre d'avoir recours à des marques de segmentation prélexicales pour localiser les frontières de mots. Cette impossibilité a cependant sa source dans la procédure utilisée pour

[49] Qui n'est pas la structure Consonne-Voyelle des mots puisque nous avons vu que le matériel de l'expérience 6 ne permettait pas de faire correspondre l'opposition liée à la fréquence des mots à une régularité de la structure phonémique.
[50] Au sens de *paradigme théorique*.

marquer des frontières entre les mots. Puisque le silence a un statut de phonème, il participe comme les autres aux processus d'activation interactive. Alors que les mots longs sont ceux qui reçoivent le plus d'activation ascendante, l'insertion d'une marque de frontière (un silence) ne suffit pas à compenser cette quantité plus importante d'activation pour les mots longs que pour leurs mots enchâssés.

Norris, McQueen, & Cutler (1995) introduisent une méthode permettant à un modèle d'utiliser des indices de segmentation prélexicaux pour influencer la reconnaissance des mots. Ils présentent les résultats de simulations informatiques dans lesquelles des informations prélexicales reposant sur les alternances prosodiques sont utilisées pour modifier les niveaux d'activation lexicale des candidats activés. La *Metrical Segmentation Strategy* (MSS) introduite par les auteurs permet de simuler l'effet des régularités accentuelles dans la reconnaissance des mots en anglais (Cutler & Norris, 1988). Le mode d'influence de la MSS dans le modèle consiste à modifier les niveaux d'activation des unités lexicales en fonction de leur alignement avec les unités phonologiques considérées comme des débuts de mots probables. En anglais en effet, une majorité de mots commence par une syllabe accentuée (ou forte). Les mots dont la syllabe initiale est non-accentuée (donc faible) sont nettement moins fréquents dans la langue. Une stratégie efficace pour localiser les frontières de mots les plus probables consiste donc à 'supposer' que les syllabes accentuées sont des débuts de mots. Pour implémenter cette 'stratégie', les auteurs établissent un lien direct entre la tâche de la MSS (localiser les syllabes accentuées) et les processus lexicaux (le niveau d'activation des unités lexicales). Les niveaux d'activation des unités lexicales sont modifiés en fonction de leur alignement avec l'unité phonologique de segmentation (l'alternance de syllabes fortes et faibles). Deux types de stratégies sont présentés. La première consiste à accroître le niveau d'activation des mots qui sont alignés avec les syllabes fortes. Lorsqu'une syllabe forte est rencontrée dans le signal, le niveau d'activation des mots qui sont alignés avec cette syllabe (donc les mots qui commencent par le phonème initial de la syllabe forte) est accru par rapport à celui des mots qui ne sont pas alignés avec la syllabe forte. Une seconde stratégie consiste au contraire à inhiber le niveau d'activation des mots qui ne sont pas alignés avec la syllabe forte. Ces deux stratégies peuvent être combinées. L'introduction de cette procédure permet aux mots qui commencent par une syllabe forte d'atteindre plus rapidement que les autres le seuil de reconnaissance. En outre, ils peuvent inhiber l'activation des autres candidats avec plus d'intensité par le biais des processus de compétition lexicale.

Pour rendre compte de l'effet des régularités phonotactiques / syllabiques, une procédure équivalente à la MSS consisterait à faire reposer les processus de segmentation sur le découpage syllabique ou phonotactique de la chaîne de phonèmes. Les mots alignés avec les débuts de syllabe subiraient un accroissement de leur niveau d'activation. Lorsqu'une séquence de phonèmes phonotactiquement illégale serait rencontrée, le niveau d'activation des candidats lexicaux qui sont alignés avec le second phonème du groupe de consonnes serait accru. Pour les stimuli dans lesquels le groupe de phonèmes médian est légal, le même processus serait appliqué mais c'est le premier phonème du groupe qui donnerait lieu à cette sur-activation. L'accroissement du niveau d'activation serait donc appliqué sur des candidats lexicaux différents en fonction de la légalité phonotactique du groupe consonantique médian. Si l'on reprend l'exemple du matériel utilisé dans les expériences de *word-spotting* présentées ici, la présentation de la séquence /bagvyʒ/ donnerait lieu à un découpage phonologique fondé sur la structuration syllabique de la séquence de phonèmes. L'activation des candidats alignés avec le début de la première syllabe (/bag/) serait accrue en raison de cet alignement entre le début de la syllabe et le début du mot. Il en irait de même pour les candidats qui sont alignés avec le début de la seconde syllabe (/vyʒ/). L'un des candidats subissant cette sur-activation au niveau de la première syllabe serait le mot 'bague'. Sur la seconde syllabe, 'vulgaire' ferait partie des candidats qui reçoivent cet accroissement de leur niveau d'activation (cf. Figure 31, gauche). On notera que ce stimulus correspond à un alignement entre frontières phonologique et lexicale. La

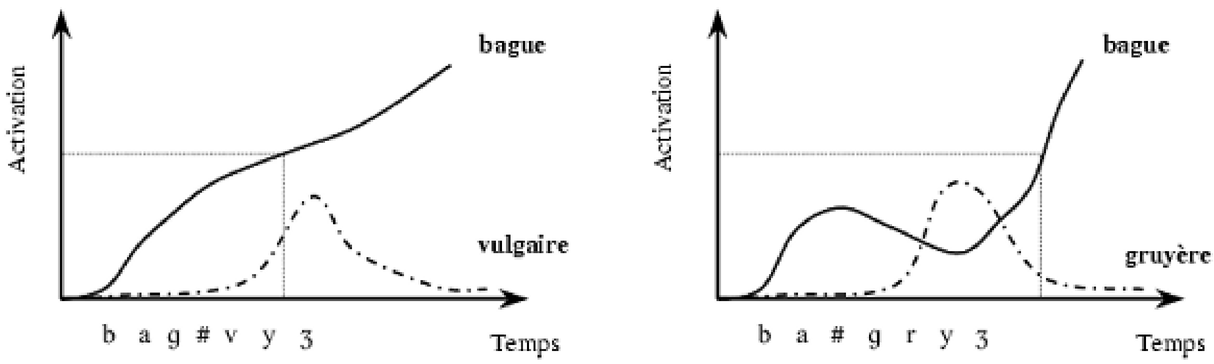

Figure 31 : Prédiction (imaginaire) du comportement d'un modèle dans lequel une procédure similaire à la Metrical Segmentation Strategy (Norris et al., 1995) accentuerait le niveau d'activation des unités lexicales qui sont alignées avec les débuts de syllabe. Le mot à reconnaître est 'bague' en position initiale d'un non-mot de deux syllabes. L'intersection de la droite horizontale avec l'axe des ordonnées indique le seuil de reconnaissance du mot. Celle de la droite verticale avec l'axe des abscisses désigne le nombre de cycles de traitement nécessaires pour atteindre le seuil. La courbe continue représente l'évolution du niveau d'activation de l'unité lexicale 'bague'. La courbe en pointillés représente l'évolution du niveau d'activation du mot aligné sur le début de la seconde syllabe ('vulgaire' à gauche et 'gruyère' à droite).

présentation de la séquence /bagryʒ/ provoquerait une segmentation syllabique différente. Frontières phonologique et lexicale ne seraient alors pas alignées. Les candidats alignés sur le début de la syllabe initiale (/ba/) seraient les mêmes que dans le précédent exemple (notamment 'bague'). Leur niveau d'activation serait accru de la même manière. La seconde syllabe (/gryʒ/) donnerait lieu à la sur-activation d'un autre ensemble de candidats parmi lesquels le mot 'gruyère' *mais pas* le mot 'rugueux'[51] (cf. Figure 31, droite). Ce dernier serait également activé mais il ne recevrait pas de sur-activation par le biais de la stratégie prélexicale. Cette sur-activation du candidat 'gruyère', lequel a le phonème /g/ en commun avec le candidat 'bague', provoquerait une inhibition plus importante du candidat 'bague' dans la séquence /bagryʒ/ que dans la séquence /bagvyʒ/. Cette inhibition plus importante du candidat 'bague' trouverait sa source dans le recouvrement des candidats qui subissent une élévation de leur niveau d'activation par le biais de la stratégie prélexicale. Lorsque 'bague' est prononcé dans la séquence /bagryʒ/, la sur-activation du candidat 'gruyère' lui permet d'inhiber les autres candidats parmi lesquels le candidat 'bague'. Celui-ci aura donc besoin d'un flux ascendant d'information plus important pour réussir à surpasser le candidat 'gruyère'. Au contraire, lorsque la séquence /bagvyʒ/ est présentée, aucun des mots du lexique qui commencent par le phonème /g/ ne subit cette sur-activation. Le mot 'gruyère' va donc être activé normalement sans accroissement provoqué. Seuls les mots commençant par la phonème /v/ vont subir cette augmentation du niveau d'activation. Mais aucun d'entre eux ne partage son phonème initial avec le candidat 'bague'. Dans cette situation d'alignement entre frontières phonologique et lexical, il n'y a pas recouvrement des candidats dont l'activation est accrue. Du fait de cette absence de recouvrement, l'unité lexicale 'bague' pourra atteindre le seuil de reconnaissance beaucoup plus rapidement, donnant ainsi lieu à l'effet d'alignement observé en *word-spotting*.

A l'éclairage de ce mode d'implémentation, l'effet d'alignement lié à la mise en œuvre de procédures de segmentation fondées sur les régularités phonotactiques / syllabiques de la langue serait déterminé par le temps nécessaire à l'unité lexicale pertinente pour atteindre le seuil de reconnaissance. Les effets observés en termes de taux d'erreur pourraient être simulés en fixant un délai au-delà duquel le modèle devrait absolument avoir choisi un candidat. L'allongement du temps nécessaire au dépassement du seuil d'activation conduirait à atteindre ce délai plus souvent lorsque les frontières syllabique et lexicale ne sont pas alignées. L'effet d'alignement

[51] Ce mot n'est en effet pas aligné avec le début de la seconde syllabe.

serait donc essentiellement déterminé par l'évolution du niveau d'activation des candidats lexicaux. D'autre part, les processus prélexicaux modifieraient directement l'état des processus lexicaux en influençant le niveau d'activation des candidats sélectionnés. L'intégration des informations en provenance des processus prélexicaux s'effectuerait par conséquent par un transfert d'informations du niveau prélexical vers le niveau de lexical ; par une modification directe des traitements effectués au niveau lexical.

3.1.2. Une proposition différente : focalisation attentionnelle sur des groupements perceptifs dérivés des ruptures dans le signal

S'il est possible que l'effet de l'accent lexical anglais se manifeste par une modification directe des niveaux d'activation lexicale, il semble que celui des alternances prosodiques du français (Banel & Bacri, 1994) ne puisse reposer sur une explication similaire ; ce qui conduit à envisager un mode d'intégration du découpage phonotactique / syllabique qui ne repose pas non plus sur ce principe. Les travaux sur les processus cognitifs impliqués dans le traitement musical et la vision incitent également à adopter une approche différente.

3.1.2.1. Intégration des informations perceptives

Alors que l'approche proposée par Norris et al. (1995) consiste à modifier directement les traitements mis en œuvre par l'étape lexicale à partir des informations fournies par les procédures de découpage prélexicales, cette proposition se doit d'être confrontée à une conception alternative de l'*intégration d'informations multiples*.

3.1.2.1.1. Le rôle des alternances prosodiques en français

En français, contrairement à l'anglais, l'accent prosodique est porté par la dernière syllabe du groupe intonatif. Les syllabes fortes ne peuvent donc pas servir à déclencher l'accès au lexique ou à accentuer le niveau d'activation des mots dont la partie initiale est alignée avec ces syllabes accentuées (Grosjean & Gee, 1987). Or Banel & Bacri (1994) montrent que des effets similaires à ceux observés par Cutler & Norris (1988) peuvent être mis en évidence en français. Les auteurs optent en faveur d'un modèle conforme aux propositions de Grosjean & Gee (1987) : les alternances prosodiques n'auraient pas pour fonction, en tout cas en français, d'accentuer le niveau d'activation des mots ou de déclencher l'accès lexical. Elles serviraient à *structurer* le signal perçu en indiquant la fin des groupes prosodiques (cf. Section 2.2.1 du Chapitre 2). *Accessoirement*, ces frontières finales pourraient faciliter la reconnaissance des mots car elles indiqueraient également une fin de mot. Il est essentiel de noter le caractère non-impérieux de l'effet de cette structuration prosodique sur la reconnaissance des mots. Il est alors possible de concevoir ces effets liés à la structure prosodique des énoncés comme dépendant de procédures de traitement qui ne seraient pas spécifiques à la langue, pas plus qu'au langage, mais reposeraient sur des procédures générales de traitement.

3.1.2.1.2. L'organisation hiérarchique des pièces musicales

Des procédures de ce type peuvent être observées dans l'organisation hiérarchique des pièces musicales (Drake, 1998). Divers niveaux d'organisation hiérarchique peuvent intervenir dans la perception de cette catégorie de stimuli. Ces niveaux correspondent à des groupements perceptifs de taille variable (cf. Figure 32). Les indices utilisés pour localiser les frontières entre les groupements -donc segmenter le signal acoustique- reposent sur les caractéristiques physiques des événements auditifs. Un changement brusque de ces caractéristiques peut déclencher la perception d'une frontière. Ces processus de segmentation effectués à divers niveaux donnent lieu à la construction d'une représentation hiérarchique de la scène musicale. Cette organisation hiérarchique peut donner lieu à des attentes sur la structure à venir de la pièce.

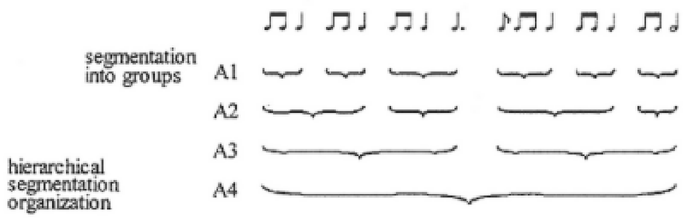

Figure 32 : Divers niveaux d'organisation hiérarchique dans la perception musicale (extrait de Drake, 1998)

Ces processus d'organisation hiérarchique des pièces musicales peuvent être mis en rapport avec les processus de segmentation du signal de parole fondés aussi bien sur les alternances prosodiques (Banel & Bacri, 1994 ; Cutler & Norris, 1988) que sur les caractéristiques des séquences de phonèmes (McQueen, 1998 ; Vroomen & de Gelder, 1999). En français, l'allongement de la durée d'une syllabe (mais aussi l'élévation de l'intensité ou de la F_0) signale la fin d'un groupe prosodique. Ces groupes prosodiques peuvent être assimilés à des groupements perceptifs correspondant à un niveau hiérarchique relativement large (2 syllabes ou plus). De même , l'occurrence d'un groupe consonantique phonotactiquement illégal permettrait de signaler une frontière de groupement à partir d'indices non plus rythmiques mais phonotactiques. Ces groupements pourraient correspondre à une seule syllabe. Parallèlement, l'occurrence d'une suite VCV induirait une segmentation syllabique fondée sur le statut d'attaque de la consonne et correspondrait également à des groupes hiérarchiques monosyllabiques. Il est essentiel de noter que des processus tout à fait similaires peuvent être décrits dans ces deux domaines d'étude. Il est même possible d'envisager que ces processus trouvent leur source dans des procédures de traitement cognitif communes et que les effets mis en évidence dans les études sur la perception de la parole peuvent reposer sur des routines de segmentation qui ne sont pas spécifiques du traitement du langage parlé (c'est ce qui est proposé par Banel & Bacri, 1994). Le modèle proposé par Cutler & Norris (1988) ne peut en tout cas pas rendre compte des effets observés en français. Puisque les groupes prosodiques français se terminent par une syllabe accentuée, il n'est pas possible de faire reposer le déclenchement des activations lexicales (ou leur modification) sur l'alignement entre syllabe accentuée et début de mot[52].

Si cette segmentation du signal de parole n'est pas spécifique du traitement du langage parlé, il n'est probablement pas nécessaire d'avoir recours à une modification directe des niveaux d'activation lexicale pour rendre compte des effets observés. Ce parallèle entre parole et musique ne suffit cependant pas à expliquer comment il serait possible de prédire les effets observés sans avoir recours à cette notion de niveau d'activation lexicale.

3.1.2.1.3. Perspectives dérivées de la théorie de la Gestalt

Les études princeps de la théorie de la forme (*Gestalt Theorie*) fournissent les fondements d'une critique de l'approche proposée par Norris et al. (1995). Les travaux de la Gestalt reposent en effet sur les principes de segmentation et groupements perceptifs des stimulations engendrées par les objets de l'environnement. Un certain nombre de *lois d'organisation perceptive* ont été proposées. Ces lois déterminent la manière dont les éléments de la scène perceptive seront organisés entre eux. Elles ont été proposées pour l'explication des processus de perception visuelle (cf. par exemple Köhler, 1929) mais peuvent également se retrouver dans l'organisation des scènes auditives (Bregman, 1990). L'un des exemples présentés par (Köhler, 1929) semble particulièrement convaincant pour se persuader de la validité d'une approche différente de celle proposée par Norris et al. (1995). Un regard rapide à la Figure 33a, ne permet d'*identifier aucune forme connue*. Au contraire, même une exposition très courte à la Figure 33b permet d'identifier le chiffre '4' dans l'image. Or le chiffre '4' est également présent dans la Figure 33a. Il est seulement dans une relation différente avec les autres composants de l'image. Dans la Figure 33c le chiffre '4' est entouré d'une ellipse. Dès que l'on a *perçu* sa présence, il devient très facile de l'isoler dans la Figure 33a. Cette différence entre la Figure 33a et la Figure 33b s'explique par la *loi de bonne continuité*. Dans la Figure 33a, les traits correspondant au chiffre '4' sont *prolongés*

[52] Il serait cependant possible d'envisager que c'est la syllabe qui suit une syllabe accentuée qui est utilisée à cet ➥

par les traits de l'image. Dans la Figure 33b au contraire, il y a une dissociation nette entre les traits du chiffre '4' -qui ne sont prolongés par aucune autre image- et ceux du reste de l'image. Ce phénomène de continuité perceptive peut donc rendre plus difficile l'*extraction* d'une information.

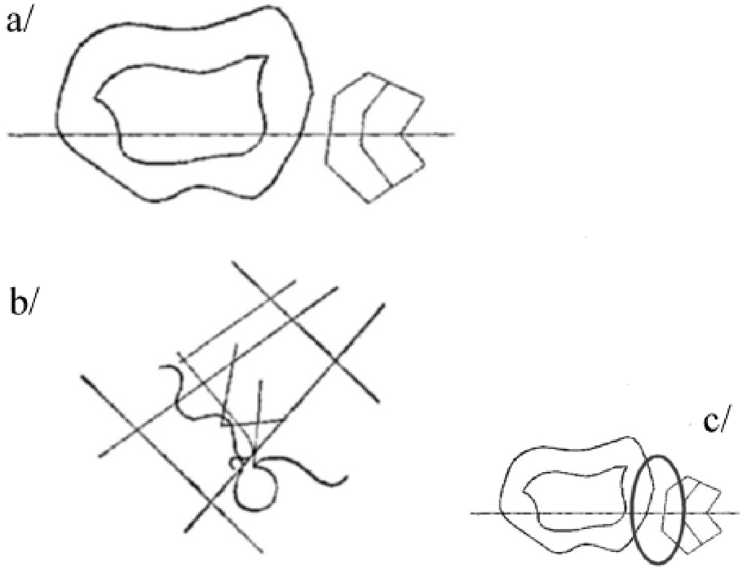

Figure 33 : Exemple de segmentation perceptive liée au phénomène de continuité des contours (ou *loi de bonne continuité*). En b/, le chiffre '4' est nettement dissocié des autres lignes. En a/, il est particulièrement difficile de percevoir ce chiffre à cause de la continuité entre les lignes du chiffre '4' et les tracés du dessin (extrait de Köhler, 1929). La figure du haut est reproduite en c/ ; le chiffre '4' y est entouré d'une ellipse afin de le localiser dans l'image.

Dans cet exemple tiré des travaux de la *Gestalt*, il n'est nullement nécessaire d'avoir recours à la notion de niveau d'activation de la représentation mentale associée au chiffre '4' pour rendre compte de l'effet perceptif. Cette difficulté à extraire le chiffre '4' de la scène visuelle est liée à des contraintes d'organisation perceptive.

3.1.2.2. Segmentation hiérarchique et niveaux d'activation lexicale

En référence aux études sur l'organisation hiérarchique des pièces musicales et aux lois d'organisation perceptive proposées par les théoriciens de la *Gestalt*, il semble qu'une approche ayant recours à des traitements cognitifs généraux peut très bien concurrencer l'approche spécifiquement langagière proposée par Norris et al. (1995). Les indices de segmentation du signal de parole seraient utilisés par des procédures d'organisation hiérarchique de la chaîne de phonèmes similaires à celles qui sont impliqués dans le traitement musical. Les effets observés sur les latences de détection ou les taux d'erreur seraient liés à des phénomènes d'intégration de multiples informations perceptives (groupements perceptifs et compétitions lexicales

---

effet.

notamment) mais aucun de ces sous-processus de traitement n'influencerait l'autre. Ces informations seraient comparées à un moment du traitement que nous ne pouvons déterminer ici -par exemple au niveau de la prise de décision- et la réponse du système dépendrait de la comparaison des deux catégories d'informations[53].

Prenons l'exemple de la détection du mot 'vague' dans les séquences /va # gryʒ/ et /vag # zyʒ/. Lorsque c'est /vag # zyʒ/ qui est présentée, les représentations phonologiques intégrées par les locuteurs permettraient de segmenter cette séquence et d'isoler les suites de phonèmes /vag/ et /zyʒ/. Cette information pourrait être transmise à une étape ultérieure du traitement. Dans le cas du stimulus /va # gryʒ/, les séquences /va/ et /gryʒ/ seraient extraites. Parallèlement à ces processus de segmentation, les séquences entières /vagryʒ/ et /vagzyʒ/ seraient traitées sans prendre en compte les indices phonologiques de segmentation et donneraient lieu à la mise en œuvre de phénomènes d'activation lexicale et de compétition entre candidats. Dans certaines situations, notamment dans celle de word-spotting, le découpage effectué par la procédure de segmentation pourrait faciliter la prise de décision des participants lorsque la segmentation effectuée correspond exactement au mot à détecter. Ainsi, quand le processus de segmentation extrait /vag/ et /zyʒ/, le mot 'vague' est directement isolé du reste du stimulus. On pourrait alors envisager que l'attention des auditeurs se focalise sur ces groupements perceptifs ; ce qui faciliterait la prise de décision et l'utilisation du produit des processus d'activation lexicale. Au contraire, la segmentation /va/ + /gryʒ/ conduirait à une focalisation attentionnelle sur des groupements perceptifs qui ne correspondent pas au mot à détecter. Les processus d'activation lexicale et de compétition entre candidats seraient alors les seuls à pouvoir fournir une information suffisante pour identifier le mot pertinent. En aucun cas, ces deux processus de traitement ne s'influenceraient mutuellement à quelque moment que ce soit. Ils se dérouleraient en parallèle et seraient intégrés ultérieurement.

Il est important de noter que selon cette proposition, et contrairement aux propositions de Norris et al. (1995 ; Norris et al., 1997), les procédures de segmentation proposées n'auraient pas pour fonction essentielle de faciliter la reconnaissance des mots. Elles serviraient plutôt à

[53] Cette discussion mériterait une réflexion plus approfondie sur deux conceptions des processus décisionnels. Un processus décisionnel peut être considéré comme simplement post-lexical et stratégique. Pour nombre de théoriciens, ce type de processus ne constitue qu'un artefact des traitements cognitifs et ne présente pas d'intérêt en soi pour la compréhension des processus cognitifs. Pour Massaro au contraire (Massaro, 1987 ; Massaro, 1989), l'intégration d'informations multiples au cours d'une étape tardive du traitement est le point essentiel de la modélisation des processus de traitement de l'information. Ces notions de *post-lexical* et *intégratif* semblent souvent correspondre au même concept analysé sous des angles différents.

structurer les signaux de parole mais reposeraient sur des procédures communes à l'ensemble des traitements cognitifs en audition. Cette structuration pourrait parfois faciliter la reconnaissance des mots mais ne serait pas l'objet primordial de ces procédures. Elle pourrait très bien ne pas intervenir dans la reconnaissance lorsque la situation expérimentale est très facile ou que les mots présentent des caractéristiques spécifiques. En outre, le modèle proposé ici serait moins sensible à l'occurrence de séquences qui ne concordent pas avec ce que recherche le système de segmentation. Quatre-vingt pour-cent des mots de l'anglais commencent par une syllabe forte, les autres mots commencent par une syllabe faible. Si l'on décide de biaiser les niveaux d'activation des mots en fonction de ces régularités, certains mots risquent de se voir inhibés alors qu'ils auraient dû être reconnus.

3.1.3. Une expérience qui permettrait de tester cette hypothèse

Afin de tester l'hypothèse proposée ici, il conviendrait de conduire une expérience dans laquelle il serait possible d'observer les effets déjà mis en évidence sans mettre les participants sous contrainte temporelle. Selon les propositions de Norris et al. (1995 ; Norris et al., 1997), c'est en dernier recours le niveau d'activation lexicale des candidats lexicaux qui détermine leur reconnaissance. Les effets d'alignement entre frontières phonologique et lexicale seraient déterminés par un allongement du temps nécessaire pour atteindre le seuil de reconnaissance. Si les participants ne doivent pas répondre le plus rapidement possible, l'activité du réseau devrait se stabiliser et permettre au mot correct d'émerger de l'ensemble des candidats quelle que soit la frontière phonologique. Selon notre proposition, il n'est pas nécessaire de mettre les participants sous contrainte temporelle pour observer cet effet. La segmentation perceptive déclenchée par les représentations phonologiques donnerait lieu à l'établissement de groupements perceptifs sur lesquels pourraient se focaliser les processus attentionnels. Toute situation difficile devrait donc permettre de faire émerger l'effet d'alignement. Nous proposons de mettre cette proposition à l'épreuve des faits en ayant recours à une tâche de *word-spotting dans du bruit*. L'insertion de bruit au niveau du seuil d'identification rendrait la tâche particulièrement difficile. Les participants disposeraient cependant d'un délai illimité pour donner leur réponse. Ils n'auraient pas besoin de répondre le plus rapidement possible. Avec un modèle dans lequel les processus de segmentation influencent directement les étapes de traitement lexicales, aucun effet ne devrait émerger. Au contraire, si la segmentation perceptive donne lieu à une focalisation attentionnelle, l'allongement des temps de réponse ne devrait pas limiter l'influence de l'effet d'alignement entre frontières phonologique et lexicale. Si les contraintes phonologiques servent à structurer le signal de parole et reposent sur des routines d'organisation perceptive non-spécifiques du traitement du langage parlé, il devrait donc être possible de mettre en évidence le même type d'effets.

Résumé

Ce dernier chapitre soulève des questions nouvelles concernant le niveau d'implémentation des procédures de segmentation lexicale fondées sur les régularités phonologiques de la langue. Si les effets n'émergent que dans certaines conditions de procédure expérimentale, il est possible que le recours à une segmentation fondée sur les régularités phonologiques ne dépendent que de la difficulté de la tâche et ne soit pas effectif pour le traitement de la parole en situation naturelle. L'absence d'effet phonologique dans une tâche de détection de phonèmes contribue à confirmer cette position. Il est cependant possible que des phénomènes acoustiques compensent l'influence des procédures de segmentation phonologique dans la tâche de détection de phonèmes. L'absence d'effet pour une partie du matériel linguistique soulève également la question du niveau d'intervention des contraintes phonologiques. Contrairement à la proposition d'une modification des niveaux d'activation lexicale par les processus de segmentation prélexicaux, un modèle d'intégration des informations dans lequel les régularités phonologiques de la langue serviraient à la structuration perceptive du signal pourrait cependant rendre compte de cette restriction sans pour autant limiter l'importance de ces procédures de segmentation. La segmentation phonologique du signal de parole serait alors utilisée pour réguler les phénomènes de focalisation attentionnelle sur les groupements perceptifs. Ces routines d'organisation perceptive se dérouleraient en parallèle avec les processus de d'activation et de compétition lexicales. Une tâche de word-spotting dans du bruit permettrait de tester cette hypothèse.

CONCLUSION

Le travail décrit dans cette thèse avait pour objet d'approfondir l'étude du rôle des régularités phonologiques de la langue dans les processus de segmentation de la parole en mots. Si les travaux précédents (McQueen, 1998 ; Vroomen & de Gelder, 1999) mettaient en évidence un effet de la structure phonologique des séquences de phonèmes sur les processus de découpage lexical du signal de parole, ils ne permettaient pas de préciser le type de connaissances mises en œuvre, pas plus que le niveau d'intervention de l'effet observé. Dans le chapitre 1, un certain nombre de travaux ont été passés en revue qui concernaient le rôle de diverses classes de connaissances dans l'identification des mots. Ceux-ci ont donné lieu à des discussions sur l'interprétation à donner des effets observés. Cette présentation a permis de mettre en évidence la difficulté à contrôler les diverses variables qui peuvent intervenir dans la reconnaissance des mots. Les études recensées dans le chapitre 2 portaient plus spécifiquement sur la problématique de la localisation des frontières lexicales en parole continue. Après avoir présenté des modèles dans lesquels la segmentation lexicale est une conséquence de l'accès au lexique, les divers indices prélexicaux de localisation des frontières ont été présentés. Nous nous sommes attardé plus longuement sur le rôle éventuel des indices phonotactiques ou syllabiques, problématique centrale de notre étude. Après une présentation des différentes approches théoriques de la structure des représentations phonologiques, le concept de *syllabe* a été abordé plus en détails et mis en relation avec la notion de *forme phonotactiquement illégale* ; ce qui conduit à poser la question de la distinction entre représentation hiérarchique vs. linéaire des séquences de phonèmes comparées dans les études expérimentales (chapitre 3). Alors que les données expérimentales observées ont été interprétées par les auteurs comme la preuve d'un recours à des connaissances sur les régularités phonologiques de la langue dans les processus de traitement du langage parlé, il a été envisagé que d'autres variables -confondues avec ces régularités phonologiques- puissent rendre compte des effets observés. Une analyse distributionnelle des groupes de consonnes dans un lexique informatisé de la langue française fait émerger une étroite relation entre la structure phonologique des séquences de phonèmes (tautosyllabique vs. hétérosyllabique ou phonotactiquement légal vs. illégal) et leur fréquence d'occurrence dans la langue (chapitre 4). Cette relation conduit à envisager les effets observés sous trois angles différents inspirés de trois modèles des processus de segmentation lexicale. Les études expérimentales des chapitres 5 et 6 ont été conduites afin d'approfondir la compréhension de ces effets et de dissocier les contributions respectives de la fréquence et de la structure phonologique des séquences de phonèmes. Les résultats obtenus confirment en partie la proposition de l'influence des régularités phonologiques dans les processus de segmentation du signal de parole en mots. Diverses observations ont cependant conduit à développer une réflexion plus approfondie sur le niveau d'implémentation des effets observés. La question du mode d'intégration des processus de segmentation phonologique du signal de parole dans un modèle du traitement du langage parlé a finalement été abordée. Un modèle dans lequel ces procédures constitueraient des routines d'organisation perceptive indépendantes des processus langagiers a été proposé en contrepoint des propositions actuelles de modélisation dans le cadre desquelles ces procédures de segmentation auraient pour fonction de modifier directement les niveaux d'activation des candidats lexicaux.

RÉFÉRENCES BIBLIOGRAPHIQUES

Aho, A. V., Kernighan, B. W., & Weinberger, P. J. (1988). *The Awk programming language*. Reading: Mass.: Addislon-Wesley Publishing.

Altmann, G. T. M. (1990). Lexical statistics and cognitive models of speech processing. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 211-235). Cambridge, Mass.: The MIT Press.

Apoux, F., Berthommier, F., Bacri, N., & Lorenzi, C. (1998). Effet du renforcement des modulations temporelles en sous-bandes sur la reconnaissance de la parole : Résultats préliminaires. *XXII^èmes^ Journées d'Etude sur la Parole*, 15-19 Juin, Martigny : Suisse.

Arnold, G. F. (1956). A phonological approach to vowel, consonant and syllable in modern French. *Lingua, 5*, 253-287.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321-324?

Autesserre, D., & Chafcouloff, M. (1999). Etude expérimentale du rôle de l'organisation syllabique dans la prédiction des variantes de /R/ en français. *IIèmes Journées d'Etudes Linguistiques*, 25-27 Mars, Nantes, France.

Banel, M.-H., & Bacri, N. (1994). On metrical patterns and lexical parsing in French. *Speech Communication, 15*, 115-126.

Banel, M.-H., & Bacri, N. (1997). Reconnaissance de la parole et indices de segmentation métriques et phonotactiques. *L'Année Psychologique, 97*, 77-112.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94*, 115-147.

Blumstein, S. E. (1986). On acoustic invariance in speech. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 178-193). Hillsdale: NJ: Lawrence Erlbaum Associates.

Boland, J. E., & Cutler, A. (1996). Interaction with autonomy: Multiple Output models and the inadequacy of the Great Divide. *Cognition, 58*, 309-320.

Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, MA: MIT Press.

Bregman, A. S., Colantonio, C., & Ahad, P. A. (1999). Is a common grouping mechanism involved in the phenomena of illusory continuity and stream segregation? *Perception & Psychophysics, 61*, 195-205.

Brent, M. R. (1996). Advances in the computational study of language acquisition. *Cognition, 61*, 1-38.

Brent, M. R. (1997). Toward a unified model of Lexical Acquisition and Lexical Access. *Journal of Psycholinguistic Research, 26*, 363-375.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61*, 93-125.

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica, 45*, 140-155.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica, 49*, 155-180.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics, 24*, 209-244.

Byrd, D., & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics, 24*, 263-282.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English* (S. P. Encrevé (1973). Principes de phonologie générative. Paris, Trad.). New York: Harper and Row.

Church, K. W. (1987). Phonological parsing and lexical retrieval. *Cognition, 25*, 53-69.

Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook, 2*, 225-252.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 283-333). Cambridge, UK: Cambridge University Press.

Clements, G. N., & Keyser, S. J. (1983). *CV Phonology: a generative theory of the syllable*. Cambridge, MA: MIT Press.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. G. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language, 37*, 463-480.

Content, A., Meunier, C., Frauenfelder, U. H., & Kearns, R. (1996). The locus of the syllable effect: prelexical or lexical? *AMLaP Conference*, Turin, Italie.

Content, A., Mousty, P., & Radeau, M. (1990). Brulex, une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique, 90*, 551-570.

Cornulier, B. d. (1986). Sur la notion de consonne et de syllabe en français. *Linguisticae Investigationes, 10*, 275-287.

Cutler, A., Butterfield, S., & Williams, J. N. (1987). The perceptual integrity of syllabic onsets. *Journal of Memory and Language, 26*, 406-418.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133-142.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*, 385-400.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology, 19*, 141-177.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merril Garrett* (pp. 113-134). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113-121.

Delgutte, B. (1987). Peripheral auditory processing of speech information: Implications from a physiological study of intensity discrimination. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception.* (pp. 333-353). Dordrecht: Netherlands: Martinus Nijhoff Publishing.

Dell, F. (1995). Consonant clusters and phonological syllables in French. *Lingua, 95*, 5-26.

Drake, C. (1998). Psychological processes involved in the temporal organization of complex auditory sequences: Universal and acquired processes. *Music Perception, 16*, 11-26.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4*, 99-109.

El-Bèze, M. (1996). Quelques apports des modèles de langage probabilistes. In H. Méloni (Ed.), *Fondements et perspectives en traitement automatique de la parole* (pp. 251-261): AUPELP - UREF.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Elman, J. L., & McClelland, J. L. (1988). Cognitive representation of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language, 27*, 143-165.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America, 97*, 1893-1904.

Flanagan, J. L. (1972). *Speech analysis, synthesis and perception*. Berlin: Springer.

Foss, D. J., & Blank, M. A. (1980). Identifying the speech codes. *Cognitive Psychology, 12*, 1-31.

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

Frauenfelder, U., & Peeters, G. (1990). Lexical segmentation in TRACE: An exercise in simulation. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 50-86). Cambridge, Mass: The MIT Press.

Frauenfelder, U. H. (1996). Models of spoken word recognition. In T. Dijskstra & K. De Smedt (Eds.), *Computational Psycholinguistics. AI and connectionnist models of human language processing.* (pp. 114-138). London: Taylor & Francis.

Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance, 16*, 77-91.

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics, 54*, 287-295.

Frisch, S. (1996). *Similarity and frequency in phonology.* PhD Dissertation, Northwestern University, Evanston, Illinois. [Adresse URL: http://ruccs.rutgers.edu/roa.html].

Frisch, S., Broe, M., & Pierrehumbert, J. (soumis). *Similarity and phonotactics in Arabic*. [Adresse URL: http://ruccs.rutgers.edu/roa.html].

Fudge, E. C. (1987). Branching structure within the syllable. *Journal of Linguistics, 23*, 359-377.

Fujimura, O. (1990). Demisyllables as sets of features. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech*. Cambridge, UK: Cambridge University Press.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 110-125.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251-279.

Goldsmith, J. A. (1990). *Autosegmental and Metrical Phonology*. Cambridge, Mass: Basil Blackwell.

Goslin, J., Content, A., & Frauenfelder, U. H. (1999). Syllable segmentation: Are humans consistent? *Eurospeech'99*, 5-10 September 1999, Budapest, Hungary. [Adresse URL: http://www.unige.ch/fapse/PSY/persons/frauenfelder/goslin/].

Graps, A. (1995). An introduction to wavelets. *IEEE Computational Science and Engineering*. [Adresse URL: http://www.amara.com].

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New-York: Wiley.

Greenberg, S., & Arai, T. (1998). Speech intelligibility is highly tolerant of cross-channel spectral asynchrony. *Joint meeting of the Acoustical Society of America and the International Congress on Acoustics*, Seattle, USA. [Adresse URL: http://www.icsi.berkeley.edu/~steveng/index.html].

Gronnum, N. (1999). Syllables at multiple levels of representation in Danish. *II^{èmes} Journées d'Etudes Linguistiques*, 25-27 Mars, Nantes, France.

Grosjean, F. (1985). The recognition of words after their acoustic offset: evidence and implications. *Perception & Psychophysics, 38*, 299-310.

Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition, 25*, 135-155.

Hallé, P. A., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: a case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception & Performance, 24*, 592-608.

Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry, 20*, 253-306.

Hooper, J. B. (1972). The syllable in phonological theory. *Language, 48*, 525-540.

Hyman, L. (1985). *A theory of phonological weight*. Dordrecht: Foris Publications.

Isel, F., & Bacri, N. (1999). Spoken-Word recognition: The access to embedded words. *Brain and Language, 68*, 61-67.

Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach* (Technical Report 8604). San Diego: University of California, Insitute for Cognitive Science.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*, 402-420.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infant's sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*, 630-645.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7*, 279-312.

Klatt, D. H. (1986). The problem of variability in speech recognition and in models of speech perception. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 300-319). Hillsdale: NJ: Lawrence Erlbaum Associates.

Klatt, D. H. (1989). Review of selected models of speech perception. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169-226). Cambridge, MA: MIT Press.

Klein, M. (1993). La syllabe comme interface de la production et de la réception phoniques. In B. Laks & M. Plénat (Eds.), *De natura sonorum : essais de phonologie* (pp. 101-142). Saint-Denis: Presses Universitaires de Vincennes.

Köhler, W. (1929). *Gestalt Psychology*. New-York: Horace Liveright.

Kucera, F., & Francis, W. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Laks, B. (1995). A connectionist account of French syllabification. *Lingua, 95*, 51-76.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

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

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics, 39*, 155-158.

Luce, P. A. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 122-147). Cambridge: Mass.: The MIT Press.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear & Hearing, 19*, 1-36.

Malécot, A. (1974). Frequency of occurrence of French phonemes and consonant clusters. *Phonetica, 29*, 158-170.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America, 69*, 548-558.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman and company.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71-102.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Lawrence Erlbaum.

Massaro, D. W. (1988). Some criticisms of connectionist models of human performance. *Journal of Memory and Language, 27*, 213-234.

Massaro, D. W. (1989). Testing between the TRACE model and the Fuzzy Logical Model of speech perception. *Cognitive Psychology, 21*, 398-421.

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics, 34*, 338-348.

Mattys, S. L. (1997). The use of time during lexical processing and segmentation: A review. *Psychonomic Bulletin & Review, 4*, 310-329.

McCarthy, J. (1988). Feature geometry and dependency: A review. *Phonetica, 43*, 84-108.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

McQueen, J. M. (1996). Word spotting. *Language and Cognitive Processes, 11*, 695-699.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*, 21-46.

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 621-638.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Language and Verbal Behavior, 20*, 298-305.

Moore, B. C. J. (1997). *An introduction to the psychology of hearing*. San Diego, CA: Academic Press.

Murre, J.-M. J., & Goebel, R. (1996). Connectionnist modelling. In T. Dijskstra & K. De Smedt (Eds.), *Computational Psycholinguistics. AI and connectionnist models of human language processing.* (pp. 49-81). London: Taylor & Francis.

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America, 62*, 714-719.

Nespor, M., & Vogel, I. (1983). Prosodic structure above the word. In A. Cutler & D. R. Ladd (Eds.), *Prosody: models and measurements* (pp. 123-140).

Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 873-889.

Norris, D. (1994). Shortlist: a connectionnist model of continuous speech recognition. *Cognition, 52*, 189-234.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 1209-1228.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299-370.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The Possible-Word Constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*, 191-243.

Norris, D. G., & Cutler, A. (1985). Juncture detection. *Linguistics, 23*, 689-705.

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of Phonetics, 18*, 375-394.

Pierrehumbert, J., & Nair, R. (1995). Word games and syllable structure. *Language and Speech, 38*, 77-114.

Pitt, M. A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics, 60*, 941-951.

Pitt, M. A., & Samuel, A. G. (1993). An empirical an meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 699-725.

Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in Generative Grammar* (Technical Report TR-2). New Brunswick, NJ: Rutgers University Cognitive Science Center.

Pulgram, E. (1965). Consonant cluster, consonant sequence, and the syllable. *Phonetica, 13*, 76-81.

Pulleyblank, D. (1989). Nonlinear phonology. *Annual Review of Anthropology, 18*, 203-226.

Rialland, A., & Badjimé, M. B. (1989). Réanalyse des tons du Bambara : des tons du nom à l'organisation générale du système. *Studies in African linguistics, 20*, 1-28.

Rossi, M. (1999). *L'intonation, le système du français*. Paris, France: Ophrys.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27-52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606-621.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General, 125*, 28-51.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology, 32*, 97-127.

Samuel, A. G., & Kat, D. (1996). Early levels of analysis of speech. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 676-694.

Segui, J., Dupoux, E., & Mehler, J. (1991). The role of the syllable in speech segmentation, phoneme identification and lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, Mass: The MIT Press.

Segui, J., Frauenfelder, U., & Mehler, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology, 72*, 471-477.

Selkirk, E. O. (1982). The syllable. In H. Van der Hulst & N. Smith (Eds.), *The structure of phonological representations (Part II)* (pp. 337-383). Dordrecht: Foris Publications.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal, 27*, 399-423.

Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*, 303-304.

Stevens, K. N. (1960). Toward a model for speech recognition. *Journal of the Acoustical Society of America, 32*, 47-55.

Stevens, K. N. (1996). Critique: Articulatory-acoustic relations and their role in speech perception. *Journal of the Acoustical Society of America, 99*, 1693-1694.

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

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America, 64*, 1358-1368.

Stevens, K. N., Keyser, S. J., & Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 426-449). Hillsdale: NJ: Lawrence Erlbaum Associates.

Summerfield, Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America, 62*, 435-448.

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language, 36*, 422-444.

Szende, T. (1980). Defining the phoneme: phenomenological aspects. *Acta Linguistica Academiae Scientiarum Hungaricae, 30*, 63-66.

Tesar, B. (1995). *Computing optimal forms in Optimality Theory: Basic syllabification*. [Adresse URL: http://ruccs.rutgers.edu/roa.html].

Titone, D., & Connine, C. M. (1997). Syllabification strategies in spoken word processing: Evidence from phonological priming. *Psychological Research, 60*, 251-263.

Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language, 27*, 87-104.

Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler & D. R. Ladd (Eds.), *Prosody: models and measurements* (pp. 53-66).

Van der Lugt, A. (1999). *From Speech to Words.* PhD Dissertation, Nimègue, Pays-Bas.

Van Son, R. J. J. H., & Pols, L. C. W. (1997). The correlation between consonant identification and the amount of acoustic consonant reduction. *Eurospeech'97*, Rhodes, Grèce.

Van Son, R. J. J. H., & Pols, L. C. W. (1999). An acoustic description of consonant reduction. *Speech Communication, 28*, 125-140.

Van Tasell, D. J. V., Soli, S. D., Kirby, V. M., & Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *Journal of the Acoustical Society of America, 77*, 1069-1077.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken work recognition. *Journal of Memory and Language, 40*, 374-408.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain & Language, 68*, 306-311.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 98-108.

Vroomen, J., & de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 710-720.

Vroomen, J., & de Gelder, B. (1999). Lexical access of resyllabified words: Evidence from phoneme monitoring. *Memory and Cognition, 27*, 413-421.

Vroomen, J., van den Bosch, A., & de Gelder, B. (1998). A connectionist model for bootstrap learning of syllabic structure. *Language and Cognitive Processes, 13*, 193-220.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*, 392-393.

Winer, B. J. (1971). *Statistical principles in experimental design*. New-York: McGraw Hill.

INDEX DES ILLUSTRATIONS

Figure 1 : Illustration du codage tonotopique effectué par la cochlée schématiquement représentée sur la gauche du graphique. Le traitement effectué peut en partie s'assimiler à une analyse en temps réel de l'enveloppe spectrale, le signal de sortie correspondant alors approximativement à ce que l'on observe sur un spectrogramme. *21*

Figure 2 : La phrase 'Le menuisier a scié une planche et l'a rabotée' ; à gauche dans sa forme originale et à droite après suppression des composantes spectrales fines. *23*

Figure 3 : Désynchronisation artificielle de la phrase 'The leagues are full of guys like that' (d'après Greenberg & Arai, 1998). Chaque spectrogramme correspond à une durée moyenne de désynchronisation de a/ 0 ms (signal original), b/ 40 ms, c/ 120 ms, d/ 220 ms. Les stimuli correspondants sont accessibles sur le site `http://www.ICSI.berkeley.edu/~steveng` *24*

Figure 4 : Schéma représentant le phénomène de coarticulation. On peut voir l'influence des voyelles sur la prononciation du phonème /g/ (d'après Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). *27*

Figure 5 : Le modèle MERGE (Norris et al., 2000). Deux catégories d'informations sont disponibles à la sortie du traitement : la prise de décision (le niveau d'activation des nœuds de la couche de droite) peut aussi bien reposer sur le niveau d'activation des nœuds phonémiques (en bas) que sur celui des nœuds lexicaux (en

haut). Le type de représentation utilisé ne permet pas de faire apparaître les boucles de récurrence. 40

Figure 6 : Exemples de stimuli utilisés dans une tâche de dévoilement progressif (ou gating). Ici, le mot utilisé pour créer une série est 'casquette'. La colonne de gauche indique approximativement les changements de phonèmes perçus. 55

Figure 7 : Représentation graphique du réseau implémenté dans TRACE (McClelland & Elman, 1986). Extrait de Frauenfelder (1996). 58

Figure 8 : Illustration d'une boucle de récurrence locale dans un réseau neuronal (extrait de Murre & Goebel, 1996). 61

Figure 9 : Représentation interne des segments dans le cadre de SPE (Chomsky & Halle, 1968). 85

Figure 10 : Représentations non-linéaires décrivant les phénomènes (a.) d'affrication, (b.) de tons montants ou descendants appliqués à un phonème unique ou (c.) de gémination. 90

Figure 11 : Représentation de la règle de dégémination du turc dans le cadre de la théorie de Clements & Keyser (1983). 93

Figure 12 : Représentation hautement hiérarchique de la syllabe en constituants phonémiques remplissant les fonctions d'attaque, noyau et coda. 98

Figure 13 : Représentation moraïque de la syllabe. 99

Figure 14 : Comparaison de la représentation de deux mots (l'un français, 'traque' ; l'autre anglais, 'tree') en fonction du choix de la structure syllabique (arborescence hautement hiérarchique en a. et c., syllabe moraïque en b. et d.). 100

Figure 15 : Enveloppe d'intensité du logatome /aba/ mettant en évidence la différence de quantité d'énergie caractérisant deux classes de phonèmes : les occlusives sonores (comme /g/) et les voyelles. 105

Figure 16 : Distribution de diverses classes de phonèmes selon l'échelle de sonorité. 105

Figure 17 : Représentation du phénomène d'affixation dans la morphologie non-concaténative de l'arabe (extrait de Frisch et al., soumis). La tire consonantique constitue la racine verbale. Les éléments de la tire vocalique sont insérés dans la racine par ancrage à la tire squelettale CV et donnent ainsi lieu à la forme verbale effective. 107

Figure 18 : Nombre moyen d'occurrences des groupes de consonnes en fonction de leur structure phonémique (mode d'articulation). Données calculées sur le corpus BRULEX (Content et al., 1990). 130

Figure 19 : Probabilité transitionnelle moyenne des groupes de consonnes en fonction de leur structure phonémique (mode d'articulation). Les fréquences du groupe de la consonne initiale

sont calculées à partir de la base de données BRULEX (Content et al., 1990). 131

Figure 20 : Probabilité d'occurrence moyenne des groupes de consonnes en fonction de leur structure phonémique (mode d'articulation). Données calculées sur le corpus BRULEX (Content et al., 1990). 132

Figure 21 : Distribution des probabilités d'occurrence pour les groupes de consonnes à initiale occlusive. 139

Figure 22 : Distribution des probabilités d'occurrence pour les groupes de consonnes à initiale fricative. 140

Figure 23 : Nombre moyen d'occurrences en début de mot. Groupes de consonnes classés en fonction de leur mode d'articulation. 145

Figure 24 : Probabilité moyenne d'occurrence en début de mot. Groupes de consonnes classés en fonction de leur mode d'articulation. 146

Figure 25 : Distribution des probabilités d'occurrence en début de mot pour les groupes de consonnes à initiale occlusive. 148

Figure 26 : Distribution des probabilités d'occurrence en position initiale de mot pour les groupes de consonnes à initiale fricative. 149

Figure 27 : Représentation graphique des interprétations reposant sur la fréquence des groupes de consonnes (effet prélexical pour une segmentation probabiliste ; effet lexical pour les phénomènes de compétitions lexicales. 159

Figure 28 : Temps de réaction moyens (en ms) observés dans les trois conditions expérimentales de contexte phonologique de l'Expérience 1. Les barres verticales représentent l'erreur-standard (SE). 170

Figure 29 : Temps de réaction moyens (en ms) observés pour les 18 mots de l'échantillon ayant le nombre de compétiteurs le plus élevé. Les barres verticales représentent l'erreur-standard (SE). 176

Figure 30 : Schéma représentant les résultats obtenus par Browman & Goldstein (1988) dans leur étude sur l'organisation articulatoire des séquences de consonnes en fonction de leur structure syllabique. La courbe du bas représente une mesure articulatoire des taux de contact entre langue et palais. 184

Figure 31 : Prédiction (imaginaire) du comportement d'un modèle dans lequel une procédure similaire à la Metrical Segmentation Strategy (Norris et al., 1995) accentuerait le niveau d'activation des unités lexicales qui sont alignées avec les débuts de syllabe. Le mot à reconnaître est 'bague' en position initiale d'un non-mot de deux syllabes. L'intersection de la droite horizontale avec l'axe des ordonnées indique le seuil de reconnaissance du mot. Celle de la droite verticale avec l'axe des abscisses désigne le nombre de cycles de traitement nécessaires pour atteindre le seuil. La courbe continue représente l'évolution du niveau d'activation de l'unité lexicale 'bague'. La courbe en pointillés représente l'évolution du

niveau d'activation du mot aligné sur le début de la seconde syllabe ('vulgaire' à gauche et 'gruyère' à droite). *250*

Figure 32 : Divers niveaux d'organisation hiérarchique dans la perception musicale (extrait de Drake, 1998) *253*

Figure 33 : Exemple de segmentation perceptive liée au phénomène de continuité des contours (ou loi de bonne continuité). En b/, le chiffre '4' est nettement dissocié des autres lignes. En a/, il est particulièrement difficile de percevoir ce chiffre à cause de la continuité entre les lignes du chiffre '4' et les tracés du dessin (extrait de Köhler, 1929). La figure du haut est reproduite en c/ ; le chiffre '4' y est entouré d'une ellipse afin de le localiser dans l'image. *255*

INDEX DES TABLEAUX

Tableau 1 : Pourcentage de mots dans le lexique anglais dont le Point d'Unicité se situe après le phonème final. Extrait des résultats de Luce (1986) pour les mots comptant entre 3 et 7 phonèmes. *54*

Tableau 2 : Statut du groupe de consonnes médian dans l'expérience de McQueen (1998) en fonction de l'alignement entre frontières phonotactique et lexicale. *78*

Tableau 3 : Seuils de probabilité des tests de Scheffé appliqués à la comparaison de probabilités d'occurrence des groupes de consonnes à initiale occlusive (a) ou fricative (b). Les comparaisons sont effectuées par catégorie de groupe en fonction du mode d'articulation des phonèmes. Les seuils de probabilité statistiquement significatifs sont retranscrits en caractères gras et italique. Les intitulés des lignes et des colonnes correspondent au second phonème de la séquence. *135*

Tableau 4 : Exemples de groupes de consonnes attestés en début de mot dans la base de données BRULEX (Content et al., 1990) mais qui seraient certainement hétérosyllabiques en position intervocalique. *142*

Tableau 5 : Seuils de probabilité des tests de Scheffé appliqués à la comparaison de probabilités d'occurrence des groupes de consonnes à initiale occlusive (a) ou fricative (b) apparaissant en position initiale de mot. Les comparaisons sont effectuées par catégorie de groupe en fonction du mode d'articulation des phonèmes. Les seuils de probabilité statistiquement significatifs sont retranscrits en caractères gras et italique. Les intitulés des *lignes et des colonnes correspondent au second phonème de la séquence*.. *147*

Tableau 6 : Statut du groupe de consonnes médian dans l'expérience de McQueen (1998) en fonction de l'alignement entre frontières phonotactique et lexicale. *153*

Tableau 7 : Conditions expérimentales de l'Expérience 1. .... *166*

Tableau 8 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 2... *181*

Tableau 9 : Conditions expérimentales de l'Expérience 3 et détail des groupes de consonnes utilisés.... *196*

Tableau 10 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 3 en fonction du type de groupe consonantique. *198*

Tableau 11 : Conditions expérimentales de l'Expérience 4 et détail des groupes de consonnes utilisés.... *202*

Tableau 12 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 4 en fonction de la fréquence du groupe consonantique. *205*

Tableau 13 : Taux d'erreurs (en %) observés dans l'Expérience 4 pour les mots ayant peu de compétiteurs en fonction de la fréquence du groupe consonantique et de la liste expérimentale................................. *207*

Tableau 14 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 5 en fonction du type de groupe consonantique. *211*

Tableau 15 : Taux d'erreurs (en %) observés dans l'Expérience 5 pour les mots ayant peu de compétiteurs en fonction du type de groupe consonantique et de la liste expérimentale....................................... *212*

Tableau 16 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 4 en fonction de la fréquence du groupe consonantique. *216*

Tableau 17 : Conditions expérimentales de l'Expérience 7 et détail des groupes de consonnes utilisés.... *233*

Tableau 18 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 7 en fonction du type de groupe consonantique. *235*

Tableau 19 : Conditions expérimentales de l'Expérience 8 et détail des groupes de consonnes utilisés.... *237*

Tableau 20 : Temps de réaction (en ms, erreur-standard entre parenthèses) et taux d'erreurs (en %) observés dans l'Expérience 8 en fonction de la fréquence du groupe consonantique. *238*

Tableau 21 : Seuils de probabilité du test de Scheffé mesurés sur les données de fréquence d'occurrence (mots à initiale occlusive). XII

Tableau 22 : Seuils de probabilité du test de Scheffé mesurés sur les données de fréquence d'occurrence (mots à initiale fricative)... XIII

Tableau 23 : Seuils de probabilité du test de Scheffé mesurés sur les données de probabilité transitionnelle (mots à initiale occlusive). XIII

Tableau 24 : Seuils de probabilité du test de Scheffé mesurés sur les données de probabilité transitionnelle (mots à initiale fricative)... XIII

Tableau 25 : Seuils de probabilité du test de Scheffé mesurés sur les données de fréquence d'occurrence (mots à initiale occlusive). XXII

Tableau 26 : Seuils de probabilité du test de Scheffé mesurés sur les données de fréquence d'occurrence (mots à initiale fricative)... XXIII

ANNEXES

Annexe 1 : Liste des groupes de consonnes (et consonnes individuelles) recherchés dans l'analyse distributionnelle présentée au Chapitre 4. Transcription IPA (International Phonetics Association) accompagnée de la transcription propre à BRULEX (Content et al., 1990).

IPA	BRULEX
b	b
bb	bb
bd	bd
bf	bf
bɡ	bg
bk	bk
bl	bl
bm	bm
bn	bn
bŋ	bN
bp	bp
br	bR
bs	bs
bʃ	bS
bt	bt
bv	bv
bʒ	bj
bz	bz
d	d
db	db
dd	dd
df	df
dɡ	dg
dk	dk
dl	dl
dm	dm
dn	dn
dŋ	dN
dp	dp
dr	dR
ds	ds
dʃ	dS
dt	dt
dv	dv
dʒ	dj
dz	dz
f	f
fb	fb
fd	fd
ff	ff
fɡ	fg
fk	fk
fl	fl
fm	fm
fn	fn
fŋ	fN
fp	fp
fr	fR
fs	fs
fʃ	fS
ft	ft
fv	fv
fʒ	fj
fz	fz

g	g	lt	lt	nt	nt
gb	gb	lv	lv	ŋt	Nt
gd	gd	lʒ	lj	nv	nv
gf	gf	lz	lz	ŋv	Nv
gg	gg	m	m	nʒ	nj
gk	gk	mb	mb	ŋʒ	Nj
gl	gl	md	md	nz	nz
gm	gm	mf	mf	ŋz	Nz
gn	gn	mg	mg	p	p
gŋ	gN	mk	mk	pb	pb
gp	gp	ml	ml	pd	pd
gr	gR	mm	mm	pf	pf
gs	gs	mn	mn	pg	pg
gʃ	gS	mŋ	mN	pk	pk
gt	gt	mp	mp	pl	pl
gv	gv	mr	mR	pm	pm
gz	gz	ms	ms	pn	pn
gʒ	gj	mʃ	mS	pŋ	pN
k	k	mt	mt	pp	pp
kb	kb	mv	mv	pr	pR
kd	kd	mz	mz	ps	ps
kf	kf	mʒ	mj	pʃ	pS
kg	kg	n	n	pt	pt
kk	kk	ŋ	N	pv	pv
kl	kl	nb	nb	pz	pz
km	km	ŋb	Nb	pʒ	pj
kn	kn	nd	nd	r	R
kŋ	kN	ŋd	Nd	rb	Rb
kp	kp	nf	nf	rd	Rd
kr	kR	ŋf	Nf	rf	Rf
ks	ks	ng	ng	rg	Rg
kʃ	kS	ŋg	Ng	rk	Rk
kt	kt	nk	nk	rl	Rl
kv	kv	ŋk	Nk	rm	Rm
kʒ	kj	nl	nl	rn	Rn
kz	kz	ŋl	Nl	rŋ	RN
l	l	nm	nm	rp	Rp
lb	lb	ŋm	Nm	ɾɾ	RR
ld	ld	nn	nn	rs	Rs
lf	lf	nŋ	nN	rʃ	RS
lg	lg	ŋn	Nn	rt	Rt
lk	lk	ŋŋ	NN	rv	Rv
ll	ll	ŋp	Np	rʒ	Rj
lm	lm	np	np	rz	Rz
ln	ln	nr	nR	s	s
lŋ	lN	ŋɾ	NR	ʃ	S
lp	lp	ns	ns	sb	sb
lr	lR	nʃ	nS	ʃb	Sb
ls	ls	ŋs	Ns	sd	sd
lʃ	lS	ŋʃ	NS	ʃd	Sd

sf	sf	tg	tg	zb	zb
ʃf	Sf	tk	tk	ʒb	jb
sg	sg	tl	tl	ʒd	jd
ʃg	Sg	tm	tm	zd	zd
sk	sk	tn	tn	ʒf	jf
ʃk	Sk	tŋ	tN	zf	zf
sl	sl	tp	tp	ʒg	jg
ʃl	Sl	tr	tR	zg	zg
sm	sm	tʃ	tS	zk	zk
ʃm	Sm	ts	ts	ʒk	jk
sn	sn	tt	tt	zl	zl
ʃn	Sn	tv	tv	ʒl	jl
sŋ	sN	tʒ	tj	zm	zm
ʃŋ	SN	tz	tz	ʒm	jm
sp	sp	v	v	ʒn	jn
ʃp	Sp	vb	vb	zn	zn
sr	sR	vd	vd	ʒŋ	jN
ʃr	SR	vf	vf	zŋ	zN
ss	ss	vg	vg	ʒp	jp
sʃ	sS	vk	vk	zp	zp
ʃs	Ss	vl	vl	ʒr	jR
ʃʃ	SS	vm	vm	zr	zR
st	st	vn	vn	ʒs	js
ʃt	St	vŋ	vN	ʒʃ	jS
ʃv	Sv	vp	vp	zs	zs
sv	sv	vr	vR	zʃ	zS
sʒ	sj	vs	vs	ʒt	jt
ʃʒ	Sj	vʃ	vS	zt	zt
sz	sz	vt	vt	ʒv	jv
ʃz	Sz	vv	vv	zv	zv
t	t	vʒ	vj	ʒʒ	jj
tb	tb	vz	vz	ʒz	jz
td	td	ʒ	j	zʒ	zj
tf	tf	z	z	zz	zz

Annexe 2 : Script en langage Awk[54] pour l'analyse distributionnelle des groupes de consonnes dans la base de données lexicale informatisée BRULEX (Content et al., 1990).

```
BEGIN{
max=306 ;
      {for (i=1 ; i<=max ; i++)
            { getline cible[i] < "clusters.txt" }
      }
```

[54] Le langage Awk est un langage interprété. Un interpréteur libre de droits (gawk) est disponible sur le site de la *Free Software Foundation* (http://www.gnu.org).

```
      {for (i=1 ; i<=max ; i++)
            { num[i]=0 }
      }
}
{ for (i=1 ; i<=max ; i++)
      { if ($2~cible[i]) { num[i]++ }
      }
}
END{
for (i in cible)
      { print cible[i],num[i] }
}
```

Annexe 3 : Le même script modifié afin de restreindre l'analyse distributionnelle aux mots pour lesquels une information sur la fréquence d'usage est disponible dans BRULEX (Content et al., 1990).

```
BEGIN{
max=306 ;
      {for (i=1 ; i<=max ; i++)
            { getline cible[i] < "clusters.txt" }
      }
      {for (i=1 ; i<=max ; i++)
            { num[i]=0 ; freq[i]=0 }
      }
}
{ for (i=1 ; i<=max ; i++)
      { if (($2~cible[i])&&($4!~"-")) { num[i]++ ; freq[i]+=$4 }
      }
}
END{
for (i in cible)
      { print cible[i],num[i],freq[i] }
}
```

Annexe 4 : Exemple de script permettant de compter le nombre d'occurrences des séquences Consonne-Voyelle dans BRULEX (Content et al., 1990).

```
BEGIN{
```

```
max=17;
      { for (i=1; i<=max; i++)
            { getline cible[i] < "consonne.txt" }
      }
      { for (i=1; i<=max; i++)
            { num[i]=0 }
      }
}

{ for (i=1; i<=max; i++)
      { if ($2~cible[i]"[aeiouyAEIOUYéèùàâêîûäëïöüÿ]")
            { number[i]++ }
      }
}

END{
for (i in cible)
      { print cible[i],num[i],freq[i] }
}
```

Annexe 5 : Résultats de l'analyse distributionnelle réalisée sans prendre en compte la position dans les mots. Les données sont triées par catégorie phonétique de groupe puis par probabilité descendante d'occurrence.

IPA	BRULEX	Type de groupe	Nombre d'occurrences	Nombre d'occurrences pour lesquelles une information sur la fréquence du mot est disponible	Fréquence cumulée des mots dans lesquels apparaît la séquence[55]	Indice de probabilité d'occurrence	Nombre d'occurrences de la consonne initiale	Probabilité transitionnelle
sf	sf	fricative-fricative	74	52	39660	63	11813	6.26^{E}-03
ʃv	Sv	fricative-fricative	9	8	20159	52	1860	4.84^{E}-03
sv	sv	fricative-fricative	20	17	5804	50	11813	1.69^{E}-03
sʒ	sj	fricative-fricative	5	4	981	36	11813	4.23^{E}-04
vs	vs	fricative-fricative	3	3	530	32	3691	8.13^{E}-04
ss	ss	fricative-fricative	4	4	122	27	11813	3.39^{E}-04
sʃ	sS	fricative-fricative	3	1	497	27	11813	2.54^{E}-04

[55] Cette valeur de fréquence est exactement celle qui est fournie dans Brulex (Content et al., 1990). Pour obtenir une estimation de la fréquence d'usage des mots sur 1 million, il faut diviser cette valeur par 100.

ʒf	jf	fricative-fricative	1	1	263	24	2916	$3.43^{E}-04$
ʃs	Ss	fricative-fricative	2	2	84	22	1860	$1.08^{E}-03$
ʃf	Sf	fricative-fricative	2	2	54	20	1860	$1.08^{E}-03$
ff	ff	fricative-fricative	0	0	0	1	4390	$0.00^{E}+00$
fʒ	fj	fricative-fricative	0	0	0	1	4390	$0.00^{E}+00$
fs	fs	fricative-fricative	2	0	0	1	4390	$4.56^{E}-04$
fʃ	fS	fricative-fricative	0	0	0	1	4390	$0.00^{E}+00$
fv	fv	fricative-fricative	0	0	0	1	4390	$0.00^{E}+00$
fz	fz	fricative-fricative	0	0	0	1	4390	$0.00^{E}+00$
ʒʒ	jj	fricative-fricative	0	0	0	1	2916	$0.00^{E}+00$
ʒs	js	fricative-fricative	0	0	0	1	2916	$0.00^{E}+00$
ʒʃ	jS	fricative-fricative	0	0	0	1	2916	$0.00^{E}+00$
ʒv	jv	fricative-fricative	0	0	0	1	2916	$0.00^{E}+00$
ʒz	jz	fricative-fricative	0	0	0	1	2916	$0.00^{E}+00$
ʃʒ	Sj	fricative-fricative	0	0	0	1	1860	$0.00^{E}+00$
ʃʃ	SS	fricative-fricative	0	0	0	1	1860	$0.00^{E}+00$
sz	sz	fricative-fricative	0	0	0	1	11813	$0.00^{E}+00$
ʃz	Sz	fricative-fricative	0	0	0	1	1860	$0.00^{E}+00$
vf	vf	fricative-fricative	0	0	0	1	3691	$0.00^{E}+00$
vʒ	vj	fricative-fricative	0	0	0	1	3691	$0.00^{E}+00$
vʃ	vS	fricative-fricative	0	0	0	1	3691	$0.00^{E}+00$
vv	vv	fricative-fricative	0	0	0	1	3691	$0.00^{E}+00$
vz	vz	fricative-fricative	0	0	0	1	3691	$0.00^{E}+00$
zf	zf	fricative-fricative	1	0	0	1	3906	$2.56^{E}-04$
zʒ	zj	fricative-fricative	0	0	0	1	3906	$0.00^{E}+00$
zs	zs	fricative-fricative	0	0	0	1	3906	$0.00^{E}+00$
zʃ	zS	fricative-fricative	0	0	0	1	3906	$0.00^{E}+00$
zv	zv	fricative-fricative	0	0	0	1	3906	$0.00^{E}+00$
zz	zz	fricative-fricative	0	0	0	1	3906	$0.00^{E}+00$
fr	fR	fricative-liquide	518	373	263542	80	4390	$1.18^{E}-01$
vr	vR	fricative-liquide	207	150	336639	77	3691	$5.61^{E}-02$
fl	fl	fricative-liquide	359	282	158102	76	4390	$8.18^{E}-02$
vl	vl	fricative-liquide	36	29	27985	59	3691	$9.75^{E}-03$
sl	sl	fricative-liquide	61	44	14654	58	11813	$5.16^{E}-03$
ʒɾ	jR	fricative-liquide	18	15	15154	54	2916	$6.17^{E}-03$
sr	sR	fricative-liquide	38	25	5469	51	11813	$3.22^{E}-03$
ʃr	SR	fricative-liquide	18	15	5255	49	1860	$9.68^{E}-03$
zr	zR	fricative-liquide	20	19	3066	48	3906	$5.12^{E}-03$
zl	zl	fricative-liquide	15	13	672	39	3906	$3.84^{E}-03$
ʃl	Sl	fricative-liquide	8	5	1079	37	1860	$4.30^{E}-03$
ʒl	jl	fricative-liquide	4	4	241	30	2916	$1.37^{E}-03$
sm	sm	fricative-nasale	601	535	160837	79	11813	$5.09^{E}-02$
zm	zm	fricative-nasale	146	143	42996	68	3906	$3.74^{E}-02$
vm	vm	fricative-nasale	74	69	67457	67	3691	$2.00^{E}-02$
vn	vn	fricative-nasale	47	34	88109	65	3691	$1.27^{E}-02$

ʒm	jm	fricative-nasale	33	23	31049	59	2916	1.13E-02
ʃm	Sm	fricative-nasale	45	40	17178	58	1860	2.42E-02
sn	sn	fricative-nasale	13	7	755	37	11813	1.10E-03
ʒn	jn	fricative-nasale	2	2	1526	35	2916	6.86E-04
fm	fm	fricative-nasale	5	4	641	34	4390	1.14E-03
ʃn	Sn	fricative-nasale	8	5	253	31	1860	4.30E-03
zn	zn	fricative-nasale	6	1	161	22	3906	1.54E-03
fn	fn	fricative-nasale	0	0	0	1	4390	0.00E+00
fŋ	fN	fricative-nasale	0	0	0	1	4390	0.00E+00
ʒŋ	jN	fricative-nasale	0	0	0	1	2916	0.00E+00
sŋ	sN	fricative-nasale	0	0	0	1	11813	0.00E+00
ʃŋ	SN	fricative-nasale	0	0	0	1	1860	0.00E+00
vŋ	vN	fricative-nasale	0	0	0	1	3691	0.00E+00
zŋ	zN	fricative-nasale	0	0	0	1	3906	0.00E+00
st	st	fricative-occlusive	2031	1372	992113	91	11813	1.72E-01
sk	sk	fricative-occlusive	559	422	524529	83	11813	4.73E-02
sp	sp	fricative-occlusive	511	407	526126	83	11813	4.33E-02
ʃt	St	fricative-occlusive	25	21	10814	54	1860	1.34E-02
vt	vt	fricative-occlusive	17	16	3698	48	3691	4.61E-03
ft	ft	fricative-occlusive	38	22	1341	45	4390	8.66E-03
sb	sb	fricative-occlusive	25	14	1330	43	11813	2.12E-03
ʒt	jt	fricative-occlusive	6	5	3492	42	2916	2.06E-03
sg	sg	fricative-occlusive	15	8	1337	40	11813	1.27E-03
vg	vg	fricative-occlusive	3	2	880	32	3691	8.13E-04
ʃk	Sk	fricative-occlusive	5	4	223	30	1860	2.69E-03
ʃp	Sp	fricative-occlusive	5	4	198	29	1860	2.69E-03
zk	zk	fricative-occlusive	2	2	442	29	3906	5.12E-04
sd	sd	fricative-occlusive	6	4	172	28	11813	5.08E-04
zt	zt	fricative-occlusive	2	2	178	26	3906	5.12E-04
vk	vk	fricative-occlusive	2	2	59	21	3691	5.42E-04
zb	zb	fricative-occlusive	3	2	67	21	3906	7.68E-04
ʒg	jg	fricative-occlusive	1	1	80	19	2916	3.43E-04
fk	fk	fricative-occlusive	4	2	33	18	4390	9.11E-04
ʃb	Sb	fricative-occlusive	2	2	16	15	1860	1.08E-03
zg	zg	fricative-occlusive	2	1	25	14	3906	5.12E-04
fb	fb	fricative-occlusive	1	1	21	13	4390	2.28E-04
fd	fd	fricative-occlusive	0	0	0	1	4390	0.00E+00
fg	fg	fricative-occlusive	0	0	0	1	4390	0.00E+00
fp	fp	fricative-occlusive	0	0	0	1	4390	0.00E+00
ʒb	jb	fricative-occlusive	0	0	0	1	2916	0.00E+00
ʒd	jd	fricative-occlusive	1	0	0	1	2916	3.43E-04
ʒk	jk	fricative-occlusive	1	0	0	1	2916	3.43E-04
ʒp	jp	fricative-occlusive	0	0	0	1	2916	0.00E+00
ʃd	Sd	fricative-occlusive	0	0	0	1	1860	0.00E+00
ʃg	Sg	fricative-occlusive	0	0	0	1	1860	0.00E+00

vb	vb	fricative-occlusive	0	0	0	1	3691	0.00E+00
vd	vd	fricative-occlusive	0	0	0	1	3691	0.00E+00
vp	vp	fricative-occlusive	0	0	0	1	3691	0.00E+00
zd	zd	fricative-occlusive	0	0	0	1	3906	0.00E+00
zp	zp	fricative-occlusive	1	0	0	1	3906	2.56E-04
s	s	fricative-voyelle	7089	5419	6.49E+06	105	11813	6.00E-01
v	v	fricative-voyelle	2463	1812	3.23E+06	98	3691	6.67E-01
f	f	fricative-voyelle	2689	1953	2.02E+06	96	4390	6.13E-01
ʒ	j	fricative-voyelle	1834	1333	1.70E+06	94	2916	6.29E-01
z	z	fricative-voyelle	2423	1777	1.56E+06	94	3906	6.20E-01
ʃ	S	fricative-voyelle	1314	961	934185	90	1860	7.06E-01
rs	Rs	liquide-fricative	502	391	564327	83	18771	2.67E-02
rv	Rv	liquide-fricative	196	134	168011	74	18771	1.04E-02
rʒ	Rj	liquide-fricative	199	160	114967	73	18771	1.06E-02
rʃ	RS	liquide-fricative	116	93	129846	71	18771	6.18E-03
rf	Rf	liquide-fricative	107	74	65387	67	18771	5.70E-03
lv	lv	liquide-fricative	80	61	40913	64	11049	7.24E-03
ls	ls	liquide-fricative	79	54	11695	58	11049	7.15E-03
lf	lf	liquide-fricative	57	44	3970	52	11049	5.16E-03
lʒ	lj	liquide-fricative	17	11	6899	49	11049	1.54E-03
lʃ	lS	liquide-fricative	12	8	1282	40	11049	1.09E-03
lz	lz	liquide-fricative	14	4	136	27	11049	1.27E-03
rz	Rz	liquide-fricative	8	2	207	26	18771	4.26E-04
rl	Rl	liquide-liquide	135	94	46591	66	18771	7.19E-03
lr	lR	liquide-liquide	31	29	9550	54	11049	2.81E-03
ɾɾ	RR	liquide-liquide	14	11	2135	44	18771	7.46E-04
ll	ll	liquide-liquide	6	5	2674	41	11049	5.43E-04
rm	Rm	liquide-nasale	564	439	530331	84	18771	3.00E-02
rn	Rn	liquide-nasale	439	315	260003	79	18771	2.34E-02
lm	lm	liquide-nasale	273	229	183004	76	11049	2.47E-02
rŋ	RN	liquide-nasale	15	13	5261	48	18771	7.99E-04
ln	ln	liquide-nasale	6	6	1257	39	11049	5.43E-04
lŋ	lN	liquide-nasale	0	0	0	1	11049	0.00E+00
rt	Rt	liquide-occlusive	737	511	673226	85	18771	3.93E-02
rd	Rd	liquide-occlusive	616	387	379239	82	18771	3.28E-02
lt	lt	liquide-occlusive	300	194	108734	73	11049	2.72E-02
rb	Rb	liquide-occlusive	283	213	78763	72	18771	1.51E-02
rk	Rk	liquide-occlusive	224	173	97383	72	18771	1.19E-02
rp	Rp	liquide-occlusive	150	135	57920	69	18771	7.99E-03
rg	Rg	liquide-occlusive	133	94	53551	67	18771	7.09E-03
lk	lk	liquide-occlusive	76	48	26000	61	11049	6.88E-03
lg	lg	liquide-occlusive	29	22	26098	58	11049	2.62E-03
lp	lp	liquide-occlusive	52	38	7634	55	11049	4.71E-03
ld	ld	liquide-occlusive	34	19	12419	54	11049	3.08E-03
lb	lb	liquide-occlusive	26	24	4047	50	11049	2.35E-03

l	l	liquide-voyelle	3852	2758	1.72E+06	97	11049	3.49E-01
r	R	liquide-voyelle	3946	2869	1.91E+06	97	18771	2.10E-01
ms	ms	nasale-fricative	11	9	376	35	7749	1.42E-03
nv	nv	nasale-fricative	4	4	253	30	6693	5.98E-04
mv	mv	nasale-fricative	3	3	118	25	7749	3.87E-04
ns	ns	nasale-fricative	6	3	93	24	6693	8.96E-04
mz	mz	nasale-fricative	3	1	89	19	7749	3.87E-04
mʒ	mj	nasale-fricative	3	1	42	16	7749	3.87E-04
nʃ	nS	nasale-fricative	4	1	12	11	6693	5.98E-04
mf	mf	nasale-fricative	0	0	0	1	7749	0.00E+00
mʃ	mS	nasale-fricative	0	0	0	1	7749	0.00E+00
nf	nf	nasale-fricative	3	2	0	1	6693	4.48E-04
ŋf	Nf	nasale-fricative	0	0	0	1	396	0.00E+00
nʒ	nj	nasale-fricative	0	0	0	1	6693	0.00E+00
ŋʒ	Nj	nasale-fricative	0	0	0	1	396	0.00E+00
ŋs	Ns	nasale-fricative	0	0	0	1	396	0.00E+00
ŋʃ	NS	nasale-fricative	0	0	0	1	396	0.00E+00
ŋv	Nv	nasale-fricative	0	0	0	1	396	0.00E+00
nz	nz	nasale-fricative	0	0	0	1	6693	0.00E+00
ŋz	Nz	nasale-fricative	0	0	0	1	396	0.00E+00
nr	nR	nasale-liquide	42	39	4330	52	6693	6.28E-03
ml	ml	nasale-liquide	19	17	2084	45	7749	2.45E-03
mr	mR	nasale-liquide	11	8	899	39	7749	1.42E-03
nl	nl	nasale-liquide	9	8	447	36	6693	1.34E-03
ŋr	NR	nasale-liquide	4	4	755	35	396	1.01E-02
ŋl	Nl	nasale-liquide	0	0	0	1	396	0.00E+00
nm	nm	nasale-nasale	98	84	53275	67	6693	1.46E-02
mn	mn	nasale-nasale	59	50	45211	64	7749	7.61E-03
ŋm	Nm	nasale-nasale	11	11	10662	51	396	2.78E-02
mm	mm	nasale-nasale	26	17	5353	50	7749	3.36E-03
nn	nn	nasale-nasale	1	1	8	9	6693	1.49E-04
mŋ	mN	nasale-nasale	0	0	0	1	7749	0.00E+00
nŋ	nN	nasale-nasale	0	0	0	1	6693	0.00E+00
ŋn	Nn	nasale-nasale	0	0	0	1	396	0.00E+00
ŋŋ	NN	nasale-nasale	0	0	0	1	396	0.00E+00
nt	nt	nasale-occlusive	29	22	2914	48	6693	4.33E-03
mt	mt	nasale-occlusive	7	5	3252	42	7749	9.03E-04
md	md	nasale-occlusive	2	1	3573	36	7749	2.58E-04
nk	nk	nasale-occlusive	5	4	671	34	6693	7.47E-04
nd	nd	nasale-occlusive	9	5	258	31	6693	1.34E-03
ŋp	Np	nasale-occlusive	2	2	131	24	396	5.05E-03
mb	mb	nasale-occlusive	3	1	51	17	7749	3.87E-04
ŋk	Nk	nasale-occlusive	1	1	8	9	396	2.53E-03
mp	mp	nasale-occlusive	2	1	4	6	7749	2.58E-04
nb	nb	nasale-occlusive	1	1	4	6	6693	1.49E-04

mg	mg	nasale-occlusive	1	0	0	1	7749	1.29^{E}-04
mk	mk	nasale-occlusive	0	0	0	1	7749	0.00^{E}+00
ŋb	Nb	nasale-occlusive	0	0	0	1	396	0.00^{E}+00
ŋd	Nd	nasale-occlusive	0	0	0	1	396	0.00^{E}+00
ng	ng	nasale-occlusive	1	0	0	1	6693	1.49^{E}-04
ŋg	Ng	nasale-occlusive	0	0	0	1	396	0.00^{E}+00
np	np	nasale-occlusive	2	0	0	1	6693	2.99^{E}-04
ŋt	Nt	nasale-occlusive	0	0	0	1	396	0.00^{E}+00
m	m	nasale-voyelle	6535	4926	4.29^{E}+06	103	7749	8.43^{E}-01
n	n	nasale-voyelle	4914	3500	3.18^{E}+06	100	6693	7.34^{E}-01
ŋ	N	nasale-voyelle	292	214	207435	76	396	7.37^{E}-01
ks	ks	occlusive-fricative	870	627	532566	85	9557	9.10^{E}-02
ps	ps	occlusive-fricative	249	185	160664	75	7454	3.34^{E}-02
gz	gz	occlusive-fricative	146	120	165251	73	2943	4.96^{E}-02
bʒ	bj	occlusive-fricative	23	22	55506	61	4644	4.95^{E}-03
dʒ	dj	occlusive-fricative	43	28	11974	55	6216	6.92^{E}-03
dv	dv	occlusive-fricative	17	15	11009	52	6216	2.73^{E}-03
tf	tf	occlusive-fricative	6	6	9412	48	14061	4.27^{E}-04
ds	ds	occlusive-fricative	4	4	11319	47	6216	6.44^{E}-04
tʃ	tS	occlusive-fricative	30	21	2183	47	14061	2.13^{E}-03
ts	ts	occlusive-fricative	29	17	2229	46	14061	2.06^{E}-03
gʒ	gj	occlusive-fricative	6	6	4016	44	2943	2.04^{E}-03
bz	bz	occlusive-fricative	2	2	4041	39	4644	4.31^{E}-04
bv	bv	occlusive-fricative	7	7	761	37	4644	1.51^{E}-03
dʃ	dS	occlusive-fricative	4	3	1318	36	6216	6.44^{E}-04
dz	dz	occlusive-fricative	14	6	223	31	6216	2.25^{E}-03
kv	kv	occlusive-fricative	2	2	466	30	9557	2.09^{E}-04
kf	kf	occlusive-fricative	5	4	143	28	9557	5.23^{E}-04
pz	pz	occlusive-fricative	3	3	186	27	7454	4.02^{E}-04
gs	gs	occlusive-fricative	4	2	156	25	2943	1.36^{E}-03
pf	pf	occlusive-fricative	6	3	53	22	7454	8.05^{E}-04
gv	gv	occlusive-fricative	1	1	102	20	2943	3.40^{E}-04
bs	bs	occlusive-fricative	2	1	76	19	4644	4.31^{E}-04
tʒ	tj	occlusive-fricative	2	1	59	18	14061	1.42^{E}-04
kʃ	kS	occlusive-fricative	1	1	29	15	9557	1.05^{E}-04
pv	pv	occlusive-fricative	1	1	29	15	7454	1.34^{E}-04
pʃ	pS	occlusive-fricative	1	1	4	6	7454	1.34^{E}-04
bf	bf	occlusive-fricative	0	0	0	1	4644	0.00^{E}+00
bʃ	bS	occlusive-fricative	0	0	0	1	4644	0.00^{E}+00
df	df	occlusive-fricative	4	0	0	1	6216	6.44^{E}-04
gf	gf	occlusive-fricative	0	0	0	1	2943	0.00^{E}+00
gʃ	gS	occlusive-fricative	0	0	0	1	2943	0.00^{E}+00
kʒ	kj	occlusive-fricative	0	0	0	1	9557	0.00^{E}+00
kz	kz	occlusive-fricative	0	0	0	1	9557	0.00^{E}+00
pʒ	pj	occlusive-fricative	0	0	0	1	7454	0.00^{E}+00

tv	tv	occlusive-fricative	1	0	0	1	14061	7.11E-05
tz	tz	occlusive-fricative	0	0	0	1	14061	0.00E+00
tr	tR	occlusive-liquide	2314	1700	2.27E+06	96	14061	1.65E-01
pr	pR	occlusive-liquide	1244	934	1.55E+06	92	7454	1.67E-01
bl	bl	occlusive-liquide	980	781	627036	87	4644	2.11E-01
kr	kR	occlusive-liquide	799	611	560923	85	9557	8.36E-02
pl	pl	occlusive-liquide	698	494	598839	85	7454	9.36E-02
br	bR	occlusive-liquide	629	501	466710	84	4644	1.35E-01
dr	dR	occlusive-liquide	400	304	792266	84	6216	6.44E-02
gr	gR	occlusive-liquide	792	609	275868	82	2943	2.69E-01
kl	kl	occlusive-liquide	436	337	274724	80	9557	4.56E-02
gl	gl	occlusive-liquide	240	186	130319	74	2943	8.15E-02
tl	tl	occlusive-liquide	47	39	8019	55	14061	3.34E-03
dl	dl	occlusive-liquide	17	14	2219	45	6216	2.73E-03
tm	tm	occlusive-nasale	130	118	82528	70	14061	9.25E-03
dm	dm	occlusive-nasale	65	62	80751	67	6216	1.05E-02
tn	tn	occlusive-nasale	43	37	90705	65	14061	3.06E-03
km	km	occlusive-nasale	115	109	22307	64	9557	1.20E-02
gm	gm	occlusive-nasale	51	42	22613	60	2943	1.73E-02
pm	pm	occlusive-nasale	14	14	12021	52	7454	1.88E-03
gn	gn	occlusive-nasale	29	23	3222	49	2943	9.85E-03
kn	kn	occlusive-nasale	42	27	2902	49	9557	4.39E-03
pn	pn	occlusive-nasale	16	10	1011	40	7454	2.15E-03
bm	bm	occlusive-nasale	6	4	552	33	4644	1.29E-03
bn	bn	occlusive-nasale	2	2	484	30	4644	4.31E-04
dn	dn	occlusive-nasale	9	3	160	27	6216	1.45E-03
bŋ	bN	occlusive-nasale	0	0	0	1	4644	0.00E+00
dŋ	dN	occlusive-nasale	0	0	0	1	6216	0.00E+00
gŋ	gN	occlusive-nasale	0	0	0	1	2943	0.00E+00
kŋ	kN	occlusive-nasale	0	0	0	1	9557	0.00E+00
pŋ	pN	occlusive-nasale	0	0	0	1	7454	0.00E+00
tŋ	tN	occlusive-nasale	0	0	0	1	14061	0.00E+00
kt	kt	occlusive-occlusive	687	484	351027	82	9557	7.19E-02
pt	pt	occlusive-occlusive	221	150	96265	72	7454	2.96E-02
tt	tt	occlusive-occlusive	5	5	4831	44	14061	3.56E-04
bd	bd	occlusive-occlusive	12	7	1393	40	4644	2.58E-03
kd	kd	occlusive-occlusive	8	5	1463	39	9557	8.37E-04
tp	tp	occlusive-occlusive	9	6	1257	39	14061	6.40E-04
tb	tb	occlusive-occlusive	9	7	772	37	14061	6.40E-04
kb	kb	occlusive-occlusive	7	3	833	34	9557	7.32E-04
dk	dk	occlusive-occlusive	5	5	278	31	6216	8.04E-04
tk	tk	occlusive-occlusive	3	3	339	30	14061	2.13E-04
pp	pp	occlusive-occlusive	1	1	204	23	7454	1.34E-04
db	db	occlusive-occlusive	5	2	71	22	6216	8.04E-04
pk	pk	occlusive-occlusive	8	3	45	21	7454	1.07E-03

gb	gb	occlusive-occlusive	3	1	85	19	2943	$1.02^{E}-03$
pg	pg	occlusive-occlusive	1	1	68	18	7454	$1.34^{E}-04$
tg	tg	occlusive-occlusive	2	1	55	17	14061	$1.42^{E}-04$
kp	kp	occlusive-occlusive	2	1	38	16	9557	$2.09^{E}-04$
td	td	occlusive-occlusive	7	1	29	15	14061	$4.98^{E}-04$
bt	bt	occlusive-occlusive	1	1	21	13	4644	$2.15^{E}-04$
dp	dp	occlusive-occlusive	7	1	12	11	6216	$1.13^{E}-03$
bb	bb	occlusive-occlusive	0	0	0	1	4644	$0.00^{E}+00$
bg	bg	occlusive-occlusive	0	0	0	1	4644	$0.00^{E}+00$
bk	bk	occlusive-occlusive	0	0	0	1	4644	$0.00^{E}+00$
bp	bp	occlusive-occlusive	0	0	0	1	4644	$0.00^{E}+00$
dd	dd	occlusive-occlusive	0	0	0	1	6216	$0.00^{E}+00$
dg	dg	occlusive-occlusive	0	0	0	1	6216	$0.00^{E}+00$
dt	dt	occlusive-occlusive	0	0	0	1	6216	$0.00^{E}+00$
gd	gd	occlusive-occlusive	0	0	0	1	2943	$0.00^{E}+00$
gg	gg	occlusive-occlusive	0	0	0	1	2943	$0.00^{E}+00$
gk	gk	occlusive-occlusive	0	0	0	1	2943	$0.00^{E}+00$
gp	gp	occlusive-occlusive	0	0	0	1	2943	$0.00^{E}+00$
gt	gt	occlusive-occlusive	0	0	0	1	2943	$0.00^{E}+00$
kg	kg	occlusive-occlusive	0	0	0	1	9557	$0.00^{E}+00$
kk	kk	occlusive-occlusive	1	0	0	1	9557	$1.05^{E}-04$
pb	pb	occlusive-occlusive	0	0	0	1	7454	$0.00^{E}+00$
pd	pd	occlusive-occlusive	0	0	0	1	7454	$0.00^{E}+00$
t	t	occlusive-voyelle	9111	6630	$6.55^{E}+06$	106	14061	$6.48^{E}-01$
d	d	occlusive-voyelle	4671	3500	$4.84^{E}+06$	102	6216	$7.51^{E}-01$
p	p	occlusive-voyelle	4399	3227	$4.39^{E}+06$	102	7454	$5.90^{E}-01$
k	k	occlusive-voyelle	4835	3468	$3.75^{E}+06$	101	9557	$5.06^{E}-01$
b	b	occlusive-voyelle	2696	2068	$1.12^{E}+06$	94	4644	$5.81^{E}-01$
g	g	occlusive-voyelle	1498	1087	704662	89	2943	$5.09^{E}-01$

Annexe 6 : Seuils de probabilité du test de Scheffé appliqué sur les données de fréquence d'occurrence et de probabilité transitionnelles obtenues sans prendre en compte la position dans les mots.

Tableau 21 : Seuils de probabilité du test de Scheffé mesurés sur les données de fréquence d'occurrence (mots à initiale occlusive).

Occlusive initiale	liquide	fricative	occlusive	nasale
voyelle	***0,000***	***0,000***	***0,000***	***0,000***
liquide		***0,036***	***0,030***	0,071
fricative			0,999	0,999
occlusive				1,000

Tableau 22 : Seuils de probabilité du test de Scheffé mesurés sur les données de fréquence d'occurrence (mots à initiale fricative).

Fricative initiale	liquide	fricative	occlusive	nasale
voyelle	***0,000***	***0,000***	***0,000***	***0,000***
liquide		0,984	0,999	0,999
fricative			0,970	0,998
occlusive				0,999

Tableau 23 : Seuils de probabilité du test de Scheffé mesurés sur les données de probabilité transitionnelle (mots à initiale occlusive).

Occlusive initiale	liquide	fricative	occlusive	nasale
voyelle	0,835	0,067	0,062	0,165
liquide		0,323	0,302	0,611
fricative			0,999	0,997
occlusive				0,996

Tableau 24 : Seuils de probabilité du test de Scheffé mesurés sur les données de probabilité transitionnelle (mots à initiale fricative).

Fricative initiale	liquide	fricative	occlusive	nasale
voyelle	***0,000***	***0,000***	***0,000***	***0,000***
liquide		0,984	0,999	0,999
fricative			***0,000***	***0,005***
occlusive				0,560

Annexe 7 : Script permettant de compter le nombre d'occurrences de chaque groupe de consonnes en début de mot.

```
BEGIN{
max=306 ;
      {for (i=1 ; i<=max ; i++)
            { getline cible[i] < "clusters.txt" }
      }
      {for (i=1 ; i<=max ; i++)
            { num[i]=0 }
      }
}
{ for (i=1 ; i<=max ; i++)
      { if ($2~"^"cible[i]) { num[i]++ }
```

```
      }
}
END{
for (i in cible)
      { print cible[i],num[i] }
}
```

Annexe 8 : Le même script modifié afin de restreindre l'analyse distributionnelle aux mots pour lesquels une information sur la fréquence d'usage est disponible dans BRULEX (Content et al., 1990).

```
BEGIN{
max=306 ;
      {for (i=1 ; i<=max ; i++)
            { getline cible[i] < "clusters.txt" }
      }
      {for (i=1 ; i<=max ; i++)
            { num[i]=0 ; freq[i]=0 }
      }
}
{ for (i=1 ; i<=max ; i++)
      { if (($2~"^"cible[i])&&($4!~"-") { num[i]++ ; freq[i]+=$4 }
      }
}
END{
for (i in cible)
      { print cible[i],num[i],freq[i] }
}
```

Annexe 9 : Exemple de script permettant de compter le nombre d'occurrences des séquences Consonne-Voyelle apparaissant en début de mot dans BRULEX (Content et al., 1990).

```
BEGIN{
max=17;
      { for (i=1; i<=max; i++)
            { getline cible[i] < "consonne.txt" }
      }
      { for (i=1; i<=max; i++)
```

```
            { num[i]=0 }
      }
}

{ for (i=1; i<=max; i++)
      { if ($2~"^"cible[i]"[aeiouyAEIOUYéèùàâêîûäëïöüÿ]")
            { number[i]++ }
      }
}

END{
for (i in cible)
      { print cible[i],num[i],freq[i] }
}
```

Annexe 10 : Résultats de l'analyse distributionnelle conduite sur les débuts de mots. Les données sont triées par catégorie phonétique de groupe puis par probabilité descendante d'occurrence.

IPA	BRULEX	Type de groupe	Nombre d'occurrences	Nombre d'occurrences pour lesquelles une information sur la fréquence du mot est disponible	Fréquence cumulée des mots dans lesquels apparaît la séquence	Indice de probabilité d'occurrence
sf	sf	fricative-fricative	5	5	2248	41
sv	sv	fricative-fricative	2	2	378	29
fʒ	fj	fricative-fricative	0	0	0	1
fs	fs	fricative-fricative	0	0	0	1
fʃ	fS	fricative-fricative	0	0	0	1
fv	fv	fricative-fricative	0	0	0	1
fz	fz	fricative-fricative	0	0	0	1
ʒf	jf	fricative-fricative	0	0	0	1
ʒs	js	fricative-fricative	0	0	0	1
ʒʃ	jS	fricative-fricative	0	0	0	1
ʒv	jv	fricative-fricative	0	0	0	1
ʒz	jz	fricative-fricative	0	0	0	1
ʃf	Sf	fricative-fricative	0	0	0	1
sʒ	sj	fricative-fricative	0	0	0	1

ʃʒ	Sj	fricative-fricative	0	0	0	1
sʃ	sS	fricative-fricative	0	0	0	1
ʃs	Ss	fricative-fricative	0	0	0	1
ʃv	Sv	fricative-fricative	0	0	0	1
sz	sz	fricative-fricative	0	0	0	1
ʃz	Sz	fricative-fricative	0	0	0	1
vf	vf	fricative-fricative	0	0	0	1
vʒ	vj	fricative-fricative	0	0	0	1
vs	vs	fricative-fricative	0	0	0	1
vʃ	vS	fricative-fricative	0	0	0	1
vz	vz	fricative-fricative	0	0	0	1
zf	zf	fricative-fricative	0	0	0	1
zʒ	zj	fricative-fricative	0	0	0	1
zs	zs	fricative-fricative	0	0	0	1
zʃ	zS	fricative-fricative	0	0	0	1
zv	zv	fricative-fricative	0	0	0	1
ff	ff	fricative-fricative	0	0	0	1
ʒʒ	jj	fricative-fricative	0	0	0	1
ss	ss	fricative-fricative	0	0	0	1
ʃʃ	SS	fricative-fricative	0	0	0	1
vv	vv	fricative-fricative	0	0	0	1
zz	zz	fricative-fricative	0	0	0	1
fr	fR	fricative-liquide	315	223	159698	76
fl	fl	fricative-liquide	172	128	61407	69
vr	vR	fricative-liquide	16	10	24313	54
sl	sl	fricative-liquide	9	4	283	31
ʃl	Sl	fricative-liquide	4	2	8	12
vl	vl	fricative-liquide	1	1	12	11
ʒl	jl	fricative-liquide	0	0	0	1
ʒr	jR	fricative-liquide	0	0	0	1
sr	sR	fricative-liquide	0	0	0	1
ʃr	SR	fricative-liquide	0	0	0	1
zl	zl	fricative-liquide	0	0	0	1
zr	zR	fricative-liquide	0	0	0	1
sn	sn	fricative-nasale	9	3	595	33
sm	sm	fricative-nasale	4	2	404	29
ʃn	Sn	fricative-nasale	3	1	21	13
fm	fm	fricative-nasale	0	0	0	1
fn	fn	fricative-nasale	0	0	0	1
fŋ	fN	fricative-nasale	0	0	0	1
ʒm	jm	fricative-nasale	6	0	0	1
ʒn	jn	fricative-nasale	0	0	0	1
ʒŋ	jN	fricative-nasale	0	0	0	1
ʃm	Sm	fricative-nasale	0	0	0	1
sŋ	sN	fricative-nasale	0	0	0	1

ʃŋ	SN	fricative-nasale	0	0	0	1
vm	vm	fricative-nasale	0	0	0	1
vn	vn	fricative-nasale	0	0	0	1
vŋ	vN	fricative-nasale	0	0	0	1
zm	zm	fricative-nasale	0	0	0	1
zn	zn	fricative-nasale	0	0	0	1
zŋ	zN	fricative-nasale	0	0	0	1
st	st	fricative-occlusive	208	151	59026	70
sp	sp	fricative-occlusive	107	85	87468	69
sk	sk	fricative-occlusive	88	68	21848	62
ft	ft	fricative-occlusive	4	1	59	18
sb	sb	fricative-occlusive	1	1	68	18
fb	fb	fricative-occlusive	0	0	0	1
fd	fd	fricative-occlusive	0	0	0	1
fg	fg	fricative-occlusive	0	0	0	1
fk	fk	fricative-occlusive	0	0	0	1
fp	fp	fricative-occlusive	0	0	0	1
ʒb	jb	fricative-occlusive	0	0	0	1
ʒd	jd	fricative-occlusive	0	0	0	1
ʒg	jg	fricative-occlusive	0	0	0	1
ʒk	jk	fricative-occlusive	0	0	0	1
ʒp	jp	fricative-occlusive	0	0	0	1
ʒt	jt	fricative-occlusive	0	0	0	1
ʃb	Sb	fricative-occlusive	0	0	0	1
sd	sd	fricative-occlusive	0	0	0	1
ʃd	Sd	fricative-occlusive	0	0	0	1
sg	sg	fricative-occlusive	0	0	0	1
ʃg	Sg	fricative-occlusive	0	0	0	1
ʃk	Sk	fricative-occlusive	0	0	0	1
ʃp	Sp	fricative-occlusive	0	0	0	1
ʃt	St	fricative-occlusive	0	0	0	1
vb	vb	fricative-occlusive	0	0	0	1
vd	vd	fricative-occlusive	0	0	0	1
vg	vg	fricative-occlusive	0	0	0	1
vk	vk	fricative-occlusive	0	0	0	1
vp	vp	fricative-occlusive	0	0	0	1
vt	vt	fricative-occlusive	0	0	0	1
zb	zb	fricative-occlusive	0	0	0	1
zd	zd	fricative-occlusive	0	0	0	1
zg	zg	fricative-occlusive	0	0	0	1
zk	zk	fricative-occlusive	0	0	0	1
zp	zp	fricative-occlusive	0	0	0	1
zt	zt	fricative-occlusive	0	0	0	1
s	s	fricative-voyelle	2095	1516	2823590	96
v	v	fricative-voyelle	833	606	1756320	90

f	f	fricative-voyelle	1127	808	1169350	90
ʒ	j	fricative-voyelle	465	330	794881	84
ʃ	S	fricative-voyelle	483	357	500025	83
z	z	fricative-voyelle	52	37	9120	55
rs	Rs	liquide-fricative	1	1	22148	43
rʒ	Rj	liquide-fricative	2	2	4432	39
lf	lf	liquide-fricative	0	0	0	1
lʒ	lj	liquide-fricative	0	0	0	1
ls	ls	liquide-fricative	0	0	0	1
lʃ	lS	liquide-fricative	0	0	0	1
lv	lv	liquide-fricative	0	0	0	1
lz	lz	liquide-fricative	0	0	0	1
rf	Rf	liquide-fricative	0	0	0	1
rʃ	RS	liquide-fricative	0	0	0	1
rv	Rv	liquide-fricative	0	0	0	1
rz	Rz	liquide-fricative	0	0	0	1
rl	Rl	liquide-liquide	6	4	12567	47
lr	lR	liquide-liquide	0	0	0	1
ll	ll	liquide-liquide	0	0	0	1
ɾɾ	RR	liquide-liquide	0	0	0	1
lm	lm	liquide-nasale	0	0	0	1
ln	ln	liquide-nasale	0	0	0	1
lŋ	lN	liquide-nasale	0	0	0	1
rm	Rm	liquide-nasale	0	0	0	1
rn	Rn	liquide-nasale	0	0	0	1
rŋ	RN	liquide-nasale	0	0	0	1
rt	Rt	liquide-occlusive	2	2	13817	44
rd	Rd	liquide-occlusive	2	1	323	25
rb	Rb	liquide-occlusive	1	1	38	16
lb	lb	liquide-occlusive	0	0	0	1
ld	ld	liquide-occlusive	0	0	0	1
lg	lg	liquide-occlusive	0	0	0	1
lk	lk	liquide-occlusive	0	0	0	1
lp	lp	liquide-occlusive	0	0	0	1
lt	lt	liquide-occlusive	0	0	0	1
rg	Rg	liquide-occlusive	0	0	0	1
rk	Rk	liquide-occlusive	0	0	0	1
rp	Rp	liquide-occlusive	0	0	0	1
r	R	liquide-voyelle	356	264	280209	79
l	l	liquide-voyelle	264	201	237276	77
mf	mf	nasale-fricative	0	0	0	1
mʒ	mj	nasale-fricative	0	0	0	1
ms	ms	nasale-fricative	0	0	0	1
mʃ	mS	nasale-fricative	0	0	0	1
mv	mv	nasale-fricative	0	0	0	1

mz	mz	nasale-fricative	0	0	0	1
nf	nf	nasale-fricative	0	0	0	1
ŋf	Nf	nasale-fricative	0	0	0	1
nʒ	nj	nasale-fricative	0	0	0	1
ŋʒ	Nj	nasale-fricative	0	0	0	1
ns	ns	nasale-fricative	0	0	0	1
nʃ	nS	nasale-fricative	0	0	0	1
ŋs	Ns	nasale-fricative	0	0	0	1
ŋʃ	NS	nasale-fricative	0	0	0	1
nv	nv	nasale-fricative	0	0	0	1
ŋv	Nv	nasale-fricative	0	0	0	1
nz	nz	nasale-fricative	0	0	0	1
ŋz	Nz	nasale-fricative	0	0	0	1
ml	ml	nasale-liquide	0	0	0	1
mr	mR	nasale-liquide	0	0	0	1
nl	nl	nasale-liquide	0	0	0	1
ŋl	Nl	nasale-liquide	0	0	0	1
nr	nR	nasale-liquide	0	0	0	1
ŋɾ	NR	nasale-liquide	0	0	0	1
mn	mn	nasale-nasale	2	2	68	21
mŋ	mN	nasale-nasale	0	0	0	1
nm	nm	nasale-nasale	0	0	0	1
ŋm	Nm	nasale-nasale	0	0	0	1
nŋ	nN	nasale-nasale	0	0	0	1
ŋn	Nn	nasale-nasale	0	0	0	1
mm	mm	nasale-nasale	0	0	0	1
nn	nn	nasale-nasale	0	0	0	1
ŋŋ	NN	nasale-nasale	0	0	0	1
mb	mb	nasale-occlusive	0	0	0	1
md	md	nasale-occlusive	0	0	0	1
mg	mg	nasale-occlusive	0	0	0	1
mk	mk	nasale-occlusive	0	0	0	1
mp	mp	nasale-occlusive	0	0	0	1
mt	mt	nasale-occlusive	0	0	0	1
nb	nb	nasale-occlusive	0	0	0	1
ŋb	Nb	nasale-occlusive	0	0	0	1
nd	nd	nasale-occlusive	0	0	0	1
ŋd	Nd	nasale-occlusive	0	0	0	1
ng	ng	nasale-occlusive	0	0	0	1
ŋg	Ng	nasale-occlusive	0	0	0	1
nk	nk	nasale-occlusive	0	0	0	1
ŋk	Nk	nasale-occlusive	0	0	0	1
np	np	nasale-occlusive	0	0	0	1
ŋp	Np	nasale-occlusive	0	0	0	1
nt	nt	nasale-occlusive	0	0	0	1

ŋt	Nt	nasale-occlusive	0	0	0	1
m	m	nasale-voyelle	2080	1384	1656060	94
n	n	nasale-voyelle	567	327	918757	85
ŋ	N	nasale-voyelle	7	2	4	9
ps	ps	occlusive-fricative	44	26	19421	57
dʒ	dj	occlusive-fricative	9	6	1277	39
ts	ts	occlusive-fricative	7	5	904	37
ks	ks	occlusive-fricative	9	2	63	21
pf	pf	occlusive-fricative	3	1	29	15
bf	bf	occlusive-fricative	0	0	0	1
bʒ	bj	occlusive-fricative	0	0	0	1
bs	bs	occlusive-fricative	0	0	0	1
bʃ	bS	occlusive-fricative	0	0	0	1
bv	bv	occlusive-fricative	0	0	0	1
bz	bz	occlusive-fricative	0	0	0	1
df	df	occlusive-fricative	0	0	0	1
ds	ds	occlusive-fricative	0	0	0	1
dʃ	dS	occlusive-fricative	0	0	0	1
dv	dv	occlusive-fricative	0	0	0	1
dz	dz	occlusive-fricative	2	0	0	1
gf	gf	occlusive-fricative	0	0	0	1
gʒ	gj	occlusive-fricative	0	0	0	1
gs	gs	occlusive-fricative	0	0	0	1
gʃ	gS	occlusive-fricative	0	0	0	1
gv	gv	occlusive-fricative	0	0	0	1
gz	gz	occlusive-fricative	0	0	0	1
kf	kf	occlusive-fricative	0	0	0	1
kʒ	kj	occlusive-fricative	0	0	0	1
kʃ	kS	occlusive-fricative	0	0	0	1
kv	kv	occlusive-fricative	0	0	0	1
kz	kz	occlusive-fricative	0	0	0	1
pʒ	pj	occlusive-fricative	0	0	0	1
pʃ	pS	occlusive-fricative	0	0	0	1
pv	pv	occlusive-fricative	0	0	0	1
pz	pz	occlusive-fricative	0	0	0	1
tf	tf	occlusive-fricative	0	0	0	1
tʒ	tj	occlusive-fricative	0	0	0	1
tʃ	tS	occlusive-fricative	3	0	0	1
tv	tv	occlusive-fricative	0	0	0	1
tz	tz	occlusive-fricative	0	0	0	1
pr	pR	occlusive-liquide	835	620	1057500	88
tr	tR	occlusive-liquide	541	432	710546	85
kr	kR	occlusive-liquide	366	274	335111	80
br	bR	occlusive-liquide	288	221	186181	76
gr	gR	occlusive-liquide	364	281	154941	76

pl	pl	occlusive-liquide	287	184	241623	76
kl	kl	occlusive-liquide	190	138	81148	70
gl	gl	occlusive-liquide	87	78	53759	66
dr	dR	occlusive-liquide	69	46	50236	64
bl	bl	occlusive-liquide	104	72	30641	63
dl	dl	occlusive-liquide	0	0	0	1
tl	tl	occlusive-liquide	0	0	0	1
pn	pn	occlusive-nasale	8	4	640	34
gn	gn	occlusive-nasale	4	4	190	29
bm	bm	occlusive-nasale	0	0	0	1
bn	bn	occlusive-nasale	0	0	0	1
bŋ	bN	occlusive-nasale	0	0	0	1
dm	dm	occlusive-nasale	0	0	0	1
dn	dn	occlusive-nasale	0	0	0	1
dŋ	dN	occlusive-nasale	0	0	0	1
gm	gm	occlusive-nasale	0	0	0	1
gŋ	gN	occlusive-nasale	0	0	0	1
km	km	occlusive-nasale	0	0	0	1
kn	kn	occlusive-nasale	1	1	0	1
kŋ	kN	occlusive-nasale	0	0	0	1
pm	pm	occlusive-nasale	0	0	0	1
pŋ	pN	occlusive-nasale	0	0	0	1
tm	tm	occlusive-nasale	0	0	0	1
tn	tn	occlusive-nasale	0	0	0	1
tŋ	tN	occlusive-nasale	0	0	0	1
bd	bd	occlusive-occlusive	0	0	0	1
bg	bg	occlusive-occlusive	0	0	0	1
bk	bk	occlusive-occlusive	0	0	0	1
bp	bp	occlusive-occlusive	0	0	0	1
bt	bt	occlusive-occlusive	0	0	0	1
db	db	occlusive-occlusive	0	0	0	1
dg	dg	occlusive-occlusive	0	0	0	1
dk	dk	occlusive-occlusive	0	0	0	1
dp	dp	occlusive-occlusive	0	0	0	1
dt	dt	occlusive-occlusive	0	0	0	1
gb	gb	occlusive-occlusive	0	0	0	1
gd	gd	occlusive-occlusive	0	0	0	1
gk	gk	occlusive-occlusive	0	0	0	1
gp	gp	occlusive-occlusive	0	0	0	1
gt	gt	occlusive-occlusive	0	0	0	1
kb	kb	occlusive-occlusive	0	0	0	1
kd	kd	occlusive-occlusive	0	0	0	1
kg	kg	occlusive-occlusive	0	0	0	1
kp	kp	occlusive-occlusive	0	0	0	1
kt	kt	occlusive-occlusive	0	0	0	1

pb	pb	occlusive-occlusive	0	0	0	1
pd	pd	occlusive-occlusive	0	0	0	1
pg	pg	occlusive-occlusive	0	0	0	1
pk	pk	occlusive-occlusive	0	0	0	1
pt	pt	occlusive-occlusive	0	0	0	1
tb	tb	occlusive-occlusive	0	0	0	1
td	td	occlusive-occlusive	0	0	0	1
tg	tg	occlusive-occlusive	0	0	0	1
tk	tk	occlusive-occlusive	0	0	0	1
tp	tp	occlusive-occlusive	0	0	0	1
bb	bb	occlusive-occlusive	0	0	0	1
dd	dd	occlusive-occlusive	0	0	0	1
gg	gg	occlusive-occlusive	0	0	0	1
kk	kk	occlusive-occlusive	0	0	0	1
pp	pp	occlusive-occlusive	0	0	0	1
tt	tt	occlusive-occlusive	0	0	0	1
d	d	occlusive-voyelle	2101	1603	3318420	97
p	p	occlusive-voyelle	2037	1406	2879190	96
k	k	occlusive-voyelle	2148	1515	2443930	96
t	t	occlusive-voyelle	1018	747	1642820	91
b	b	occlusive-voyelle	1227	938	598654	87
g	g	occlusive-voyelle	516	375	230462	79

Annexe 11 : Seuils de probabilité du test de Scheffé appliqué sur les données de fréquence d'occurrence en début de mot.

Tableau 25 : Seuils de probabilité du test de Scheffé mesurés sur les données de fréquence d'occurrence (mots à initiale occlusive).

Occlusive initiale	liquide	fricative	occlusive	nasale
voyelle	***0,000***	***0,000***	***0,000***	***0,000***
liquide		0,771	0,765	0,833
fricative			1,000	1,000
occlusive				1,000

Tableau 26 : Seuils de probabilité du test de Scheffé mesurés sur les données de fréquence d'occurrence (mots à initiale fricative).

Fricative initiale	liquide	fricative	occlusive	nasale
voyelle	***0,000***	***0,000***	***0,000***	***0,000***
liquide		0,999	0,999	0,999
fricative			0,999	1,000
occlusive				0,999

Annexe 12 : Consigne lue aux sujets pour les tâches de *word-spotting*.

Vous allez entendre des séquences de parole sans signification. Certaines d'entre elles commencent par un mot de la langue. Votre tâche consiste à détecter ces mots le plus rapidement possible. Lorsque vous identifiez un mot en position initiale de non-mot, vous devez appuyer le plus rapidement possible sur le bouton-réponse. Chaque fois que vous détectez un mot, vous devez le prononcer à voix haute après avoir appuyé sur le bouton-réponse.

Vous devez donc donner deux types de réponses : (1) appuyer le plus vite possible sur le bouton-réponse lorsque vous détectez un mot en position initiale de non-mot et (2) si vous détectez un mot, le prononcer à voix haute.

Annexe 13 : Matériel utilisé dans l'Expérience 1.

La liste des items de remplissage est présentée en Annexe 20.

Mot	Structure CV	Fréquence / 1 million	Nombre de compétiteurs du mot	Condition de contexte phonologique	Stimulus	Groupe de consonnes médian	Durée du mot initial (jusqu'à l'occlusion)	Probabilité d'occurrence brute du groupe de consonnes médian
monde	CVC	1023	6	occl.-liq.	mɔ̃dran	dr	238	85
route	CVC	240	6	occl.-liq.	rutryv	tr	260	100
faute	CVC	148	6	occl.-liq.	fotrig	tr	308	100
groupe	CCVC	140	4	occl.-liq.	grupliʒ	pl	308	91
type	CVC	136	8	occl.-liq.	tiprin	pr	179	93
jambe	CVC	114	4	occl.-liq.	ʒɑ̃blyz	bl	306	89
langue	CVC	98	16	occl.-liq.	lɑ̃gryʒ	gr	245	88
sac	CVC	60	30	occl.-liq.	saklyz	kl	284	81
date	CVC	52	6	occl.-liq.	datryʒ	tr	260	100

pente	CVC	52	17	occl.-liq.	pɑ̃trik	tr	190	100
bande	CVC	42	9	occl.-liq.	bɑ̃driʒ	dr	241	85
tombe	CVC	38	10	occl.-liq.	tɔ̃briʒ	br	243	84
guide	CVC	28	3	occl.-liq.	gidryv	dr	203	85
bombe	CVC	26	11	occl.-liq.	bɔ̃blab	bl	247	89
lac	CVC	26	21	occl.-liq.	laklyv	kr	285	87
banque	CVC	25	11	occl.-liq.	bɑ̃klab	kl	269	81
jupe	CVC	22	1	occl.-liq.	ʒyplab	pl	293	91
brute	CCVC	19	5	occl.-liq.	brytrin	tr	280	100
brique	CCVC	17	9	occl.-liq.	brikrag	kr	243	87
plaque	CCVC	17	5	occl.-liq.	plaklum	kl	234	81
stade	CCVC	16	0	occl.-liq.	stadrif	dr	301	85
pompe	CVC	14	12	occl.-liq.	pɔ̃pryv	pr	197	93
rampe	CVC	13	22	occl.-liq.	rɑ̃plun	pl	277	91
bague	CVC	12	11	occl.-liq.	baglym	gl	249	75
blague	CCVC	12	3	occl.-liq.	blagrɔ̃k	gr	295	88
grappe	CCVC	11	3	occl.-liq.	grapriʒ	pr	313	93
glande	CCVC	9	0	occl.-liq.	glɑ̃drun	dr	282	85
tube	CVC	9	12	occl.-liq.	tybrin	br	177	84
flaque	CCVC	8	1	occl.-liq.	flakrib	kr	321	87
cube	CVC	7	9	occl.-liq.	kyblan	bl	207	89
fugue	CVC	6	3	occl.-liq.	fygliʒ	gl	299	75
grippe	CCVC	6	4	occl.-liq.	griplɛ̃d	pl	328	91
trappe	CCVC	6	7	occl.-liq.	trapriz	pr	280	93
crabe	CCVC	4	0	occl.-liq.	krabryv	br	287	84
monde	CVC	1023	6	occl.-fric.	mɔ̃dzan	dz	278	43
route	CVC	240	6	occl.-fric.	rutsyv	ts	247	45
faute	CVC	148	6	occl.-fric.	fotʃig	tʃ	293	50
groupe	CCVC	140	4	occl.-fric.	grupsiʒ	ps	297	77
type	CVC	136	8	occl.-fric.	tipsin	ps	175	77
jambe	CVC	114	4	occl.-fric.	ʒɑ̃bvyz	bv	303	37
langue	CVC	98	16	occl.-fric.	lɑ̃gzyʒ	gz	259	74
sac	CVC	60	30	occl.-fric.	sakʃyz	kʃ	307	15
date	CVC	52	6	occl.-fric.	datfyʒ	tf	225	42
pente	CVC	52	17	occl.-fric.	pɑ̃tfib	tf	199	42
bande	CVC	42	9	occl.-fric.	bɑ̃dviʒ	dv	266	50
tombe	CVC	38	10	occl.-fric.	tɔ̃bziʒ	bz	225	39
guide	CVC	28	3	occl.-fric.	gidzyv	dz	234	43
bombe	CVC	26	11	occl.-fric.	bɔ̃bvab	bv	271	37
lac	CVC	26	21	occl.-fric.	lakfyv	ks	203	87
banque	CVC	25	11	occl.-fric.	bɑ̃kfab	kf	261	26
jupe	CVC	22	1	occl.-fric.	ʒypfab	pf	265	15
brute	CCVC	19	5	occl.-fric.	brytsin	ts	281	45
brique	CCVC	17	9	occl.-fric.	brikfag	kf	287	26
plaque	CCVC	17	5	occl.-fric.	plakʃum	kʃ	230	15

stade	CCVC	16	0	occl.-fric.	stadʒif	dʒ	346	57
pompe	CVC	14	12	occl.-fric.	pɔ̃pfyv	pf	217	15
rampe	CVC	13	22	occl.-fric.	rɑ̃psun	ps	280	77
bague	CVC	12	11	occl.-fric.	bagʒym	gʒ	248	43
blague	CCVC	12	3	occl.-fric.	blagzɔ̃k	gz	267	74
grappe	CCVC	11	3	occl.-fric.	grapsiʒ	ps	294	77
glande	CCVC	9	0	occl.-fric.	glɑ̃dvun	dv	292	50
tube	CVC	9	12	occl.-fric.	tybʒin	bʒ	187	60
flaque	CCVC	8	1	occl.-fric.	flakʃib	kʃ	309	15
cube	CVC	7	9	occl.-fric.	kybzan	bz	193	39
fugue	CVC	6	3	occl.-fric.	fygziʒ	gz	298	74
grippe	CCVC	6	4	occl.-fric.	gripfɛ̃d	pf	287	15
trappe	CCVC	6	7	occl.-fric.	trapfiz	pf	282	15
crabe	CCVC	4	0	occl.-fric.	krabʒyv	bʒ	288	60
monde	CVC	1023	6	occl.-occl.	mɔ̃dban	db	252	17
route	CVC	240	6	occl.-occl.	rutpyv	tp	269	22
faute	CVC	148	6	occl.-occl.	fotkig	tk	281	25
groupe	CCVC	140	4	occl.-occl.	grupkiʒ	pk	247	32
type	CVC	136	8	occl.-occl.	tipkin	pk	161	32
jambe	CVC	114	4	occl.-occl.	ʒɑ̃bdyz	bd	310	43
langue	CVC	98	16	occl.-occl.	lɑ̃gbyʒ	gb	260	25
sac	CVC	60	30	occl.-occl.	sakpyz	kp	269	16
date	CVC	52	6	occl.-occl.	datkyʒ	tk	247	25
pente	CVC	52	17	occl.-occl.	pɑ̃tkib	tk	186	25
bande	CVC	42	9	occl.-occl.	bɑ̃dbiʒ	db	230	17
tombe	CVC	38	10	occl.-occl.	tɔ̃bdiʒ	bd	205	43
guide	CVC	28	3	occl.-occl.	gidbyv	db	214	17
bombe	CVC	26	11	occl.-occl.	bɔ̃bdab	bd	265	43
lac	CVC	26	21	occl.-occl.	lakpyv	kp	239	16
banque	CVC	25	11	occl.-occl.	bɑ̃ktab	kt	212	84
jupe	CVC	22	1	occl.-occl.	ʒyptab	pt	231	73
brute	CCVC	19	5	occl.-occl.	brytkin	tk	269	25
brique	CCVC	17	9	occl.-occl.	brikpag	kp	259	16
plaque	CCVC	17	5	occl.-occl.	plaktum	kt	237	84
stade	CCVC	16	0	occl.-occl.	stadbif	db	403	17
pompe	CVC	14	12	occl.-occl.	pɔ̃ptyv	pt	194	73
rampe	CVC	13	22	occl.-occl.	rɑ̃pkun	pk	235	32
bague	CVC	12	11	occl.-occl.	bagbym	gb	244	25
blague	CCVC	12	3	occl.-occl.	blagbɔ̃k	gb	304	25
grappe	CCVC	11	3	occl.-occl.	grapkiʒ	pk	270	32
glande	CCVC	9	0	occl.-occl.	glɑ̃dbun	db	287	17
tube	CVC	9	12	occl.-occl.	tybdin	bd	178	43
flaque	CCVC	8	1	occl.-occl.	flaktib	kt	322	84
cube	CVC	7	9	occl.-occl.	kybdan	bd	200	43
fugue	CVC	6	3	occl.-occl.	fygbiʒ	gb	280	25

grippe	CCVC	6	4	occl.-occl.	griptɛ̃d	pt	268	73
trappe	CCVC	6	7	occl.-occl.	trapkiz	pk	244	32
crabe	CCVC	4	0	occl.-occl.	krabdyv	bd	287	43

Annexe 14 : Matériel utilisé dans l'Expérience 3.

La liste des items de remplissage est présentée en Annexe 21.

Mot	Structure CV	Fréquence / 1 million	Nombre de compétiteurs du mot	Condition de contexte phonologique	Transcription phonémique	Groupe de consonnes médian	Durée du mot initial (en ms. jusqu'à l'occlusion)	Probabilité d'occurrence brute	Probabilité d'occurrence positionnelle
cinq	CVC	219	13	tautosyll.	sɛ̃kliz	kl	312	80	70
longue	CVC	214	5	tautosyll.	lɔ̃glyv	gl	287	74	66
langue	CVC	98	16	tautosyll.	lɑ̃glid	gl	275	74	66
sac	CVC	60	28	tautosyll.	saklɔd	kl	327	80	70
duc	CVC	52	6	tautosyll.	dyklab	kl	248	80	70
nuque	CVC	26	2	tautosyll.	nykliz	kl	227	80	70
banque	CVC	25	11	tautosyll.	bɑ̃kran	kr	270	85	80
brique	CCVC	17	10	tautosyll.	brikrum	kr	244	85	80
truc	CCVC	17	7	tautosyll.	tryklid	kl	290	80	70
plaque	CCVC	17	5	tautosyll.	plakriʒ	kr	248	85	80
blague	CCVC	12	3	tautosyll.	blaglym	gl	312	74	66
bague	CVC	12	11	tautosyll.	baglyr	gl	279	74	66
digue	CVC	10	1	tautosyll.	diglan	gl	265	74	66
ligue	CVC	9	7	tautosyll.	liglɔg	gl	251	74	66
flaque	CCVC	8	1	tautosyll.	flakryv	kr	301	85	80
claque	CCVC	6	10	tautosyll.	klaklym	kl	284	80	70
fugue	CVC	6	3	tautosyll.	fygliʒ	gl	312	74	66
bac	CVC	5	10	tautosyll.	bakrun	kr	237	85	80
fougue	CVC	5	4	tautosyll.	fuglab	gl	294	74	66
figue	CVC	3	10	tautosyll.	figlɛʒ	gl	302	74	66
cinq	CVC	219	13	hétérosyll.	sɛ̃ktiz	kt	283	82	1
longue	CVC	214	5	hétérosyll.	lɔ̃gzyv	gz	271	66	1
langue	CVC	98	16	hétérosyll.	lɑ̃gzid	gz	266	66	1
sac	CVC	60	28	hétérosyll.	saktɔd	kt	273	82	1
duc	CVC	52	6	hétérosyll.	dyktab	kt	205	82	1
nuque	CVC	26	2	hétérosyll.	nyktiz	kt	203	82	1
banque	CVC	25	11	hétérosyll.	bɑ̃ksan	ks	262	85	21
brique	CCVC	17	10	hétérosyll.	briksum	ks	285	85	21
truc	CCVC	17	7	hétérosyll.	tryktid	kt	231	82	1

plaque	CCVC	17	5	hétérosyll.	plaksiʒ	ks	217	85	21
bague	CVC	12	11	hétérosyll.	bagzyr	gz	250	66	1
blague	CCVC	12	3	hétérosyll.	blagzym	gz	281	66	1
digue	CVC	10	1	hétérosyll.	digzan	gz	284	66	1
ligue	CVC	9	7	hétérosyll.	ligzɔg	gz	216	66	1
flaque	CCVC	8	1	hétérosyll.	flaksyv	ks	277	85	21
fugue	CVC	6	3	hétérosyll.	fygziʒ	gz	276	66	1
claque	CCVC	6	10	hétérosyll.	klaktym	kt	247	82	1
fougue	CVC	5	4	hétérosyll.	fugzab	gz	232	66	1
bac	CVC	5	10	hétérosyll.	baksun	ks	212	85	21
figue	CVC	3	10	hétérosyll.	figzɛʒ	gz	241	66	1

Annexe 15 : Matériel utilisé dans l'Expérience 4.

La liste des items de remplissage est présentée en Annexe 22.

Mot	Structure CV	Fréquence / 1 million	Nombre de compétiteurs du mot	Condition de contexte	Stimulus	Groupe de consonnes médian	Durée du mot initial (jusqu'à l'occlusion)	Probabilité d'occurrence brute	Probabilité d'occurrence positionnelle
langue	CVC	98	16	fréquent	lɑ̃gzyl	gz	313	73	1
bague	CVC	12	11	fréquent	bagzɛp	gz	287	73	1
figue	CVC	3	10	fréquent	figzab	gz	311	73	1
ligue	CVC	9	7	fréquent	ligzym	gz	254	73	1
longue	CVC	214	5	fréquent	lɔ̃gzyk	gz	286	73	1
fougue	CVC	5	4	fréquent	fugzod	gz	280	73	1
blague	CCVC	12	3	fréquent	blagzim	gz	306	73	1
fugue	CVC	6	3	fréquent	fygzab	gz	291	73	1
digue	CVC	10	2	fréquent	digzyn	gz	251	73	1
tempe	CVC	19	29	fréquent	tɑ̃psid	ps	216	75	57
coupe	CVC	25	19	fréquent	kupsan	ps	186	75	57
soupe	CVC	25	16	fréquent	suptyp	pt	315	72	1
pompe	CVC	14	12	fréquent	pɔ̃ptam	pt	198	72	1
lampe	CVC	72	9	fréquent	lɑ̃psym	ps	296	75	57
nappe	CVC	22	9	fréquent	napsɛd	ps	281	75	57
type	CVC	136	8	fréquent	tipsum	ps	201	75	57
croupe	CCVC	9	7	fréquent	kʀuptad	pt	256	72	1
pipe	CVC	27	7	fréquent	piptɔf	pt	174	72	1
trappe	CCVC	6	7	fréquent	traptyʒ	pt	252	72	1
trompe	CCVC	8	7	fréquent	trɔ̃psim	ps	277	75	57
grippe	CCVC	6	5	fréquent	gripsaf	ps	304	75	57

groupe	CCVC	140	4	fréquent	gruptan	pt	307	72	1
loupe	CVC	4	4	fréquent	luptɛn	pt	235	72	1
trempe	CCVC	3	4	fréquent	trãpsaz	ps	266	75	57
grappe	CCVC	11	3	fréquent	grapsil	ps	337	75	57
crampe	CCVC	2	2	fréquent	kraptɔf	pt	279	72	1
troupe	CCVC	113	2	fréquent	truptin	pt	262	72	1
jupe	CVC	22	1	fréquent	ʒypsim	ps	286	75	57
rampe	CVC	13	0	fréquent	rãptad	pt	282	72	1
langue	CVC	98	16	rare	lãgʒyl	gʒ	285	44	1
bague	CVC	12	11	rare	bagʒɛp	gʒ	246	44	1
figue	CVC	3	10	rare	figʒab	gʒ	262	44	1
ligue	CVC	9	7	rare	ligʒym	gʒ	261	44	1
longue	CVC	214	5	rare	lɔ̃gʒyk	gʒ	282	44	1
fougue	CVC	5	4	rare	fugʒod	gʒ	298	44	1
blague	CCVC	12	3	rare	blagʒim	gʒ	313	44	1
fugue	CVC	6	3	rare	fygʒab	gʒ	274	44	1
digue	CVC	10	2	rare	digʒyn	gʒ	232	44	1
tempe	CVC	19	29	rare	tãpʃid	pʃ	208	6	1
coupe	CVC	25	19	rare	kupʃan	pʃ	175	6	1
soupe	CVC	25	16	rare	supkyp	pk	306	21	1
pompe	CVC	14	12	rare	pɔ̃pkam	pk	190	21	1
lampe	CVC	72	9	rare	lãpʃym	pʃ	280	6	1
nappe	CVC	22	9	rare	napʃɛd	pʃ	267	6	1
type	CVC	136	8	rare	tipʃum	pʃ	202	6	1
croupe	CCVC	9	7	rare	krupkad	pk	259	21	1
pipe	CVC	27	7	rare	pipkɔf	pk	166	21	1
trappe	CCVC	6	7	rare	trapkyʒ	pk	256	21	1
trompe	CCVC	8	7	rare	trɔ̃pʃim	pʃ	269	6	1
grippe	CCVC	6	5	rare	gripʃaf	pʃ	312	6	1
groupe	CCVC	140	4	rare	grupkan	pk	304	21	1
loupe	CVC	4	4	rare	lupkɛn	pk	270	21	1
trempe	CCVC	3	4	rare	trãpʃaz	pʃ	285	6	1
grappe	CCVC	11	3	rare	grapʃil	pʃ	307	6	1
crampe	CCVC	2	2	rare	krapkɔf	pk	264	21	1
troupe	CCVC	113	2	rare	trupkin	pk	269	21	1
jupe	CVC	22	1	rare	ʒypʃim	pʃ	305	6	1
rampe	CVC	13	0	rare	rãpkad	pk	256	21	1

Annexe 16 : Matériel utilisé dans l’Expérience 5.

Les stimuli-test sont identiques à ceux de l’expérience 3. Seuls les items de remplissage supplémentaires diffèrent. Ceux-ci sont présentés en Annexe 23.

Annexe 17 : Matériel utilisé dans l'Expérience 6.

Les stimuli-test et les items de remplissage sont identiques à ceux de l'expérience 3.

Annexe 18 : Matériel utilisé dans l'Expérience 7.

La liste des items de remplissage et des 'attrapes' est présentée en Annexe 24.

Déterminant	Nom Commun	Adjectif	Structure CV	Fréquence / 1 million	Nombre de compétiteurs du mot	Condition de contexte phonologique	Groupe de consonnes médian	Durée du mot initial (jusqu'à l'occlusion)	Probabilité d'occurrence brute	Probabilité d'occurrence positionnelle
Un	sac	lavable	CVC	60	28	tautosyll.	kl	376	80	70
Un	cinq	latin	CVC	219	13	tautosyll.	kl	378	80	70
Une	banque	rivale	CVC	25	11	tautosyll.	kr	372	85	80
Un	bac	réussi	CVC	5	10	tautosyll.	kr	329	85	80
Une	brique	robuste	CCVC	17	10	tautosyll.	kr	392	85	80
Un	truc	léger	CCVC	17	7	tautosyll.	kl	365	80	70
Un	duc	lucide	CVC	52	6	tautosyll.	kl	314	80	70
Une	plaque	romaine	CCVC	17	5	tautosyll.	kr	379	85	80
Un	sac	tissé	CVC	60	28	hétérosyll.	kt	401	82	1
Un	cinq	tordu	CVC	219	13	hétérosyll.	kt	388	82	1
Une	banque	sérieuse	CVC	25	11	hétérosyll.	ks	365	85	21
Un	bac	simplifié	CVC	5	10	hétérosyll.	ks	344	85	21
Une	brique	saillante	CCVC	17	10	hétérosyll.	ks	414	85	21
Un	truc	touchant	CCVC	17	7	hétérosyll.	kt	385	82	1
Un	duc	timide	CVC	52	6	hétérosyll.	kt	325	82	1
Une	plaque	solide	CCVC	17	5	hétérosyll.	ks	369	85	21

Annexe 19 : Matériel utilisé dans l'Expérience 8.

La liste des items de remplissage et des 'attrapes' est présentée en Annexe 24.

Déterminant	Nom Commun	Adjectif	Structure CV	Fréquence / 1 million	Nombre de compétiteurs du mot	Condition de contexte	Groupe de consonnes médian	Durée du mot initial (jusqu'à l'occlusion)	Probabilité d'occurrence brute	Probabilité d'occurrence positionnelle
Une	coupe	soldée	CVC	25	19	fréquent	ps	360	75	57
Une	soupe	toxique	CVC	25	16	fréquent	pt	358	72	1
Une	lampe	sublime	CVC	72	9	fréquent	ps	374	75	57
Une	nappe	soyeuse	CVC	22	9	fréquent	ps	418	75	57
Un	type	célèbre	CVC	136	8	fréquent	ps	351	75	57
Une	croupe	tigrée	CCVC	9	7	fréquent	pt	403	72	1
Une	trompe	sinueuse	CCVC	8	7	fréquent	ps	392	75	57
Une	grippe	sournoise	CCVC	6	5	fréquent	ps	380	75	57
Un	groupe	témoin	CCVC	140	4	fréquent	pt	364	72	1
Une	grappe	sucrée	CCVC	11	3	fréquent	ps	427	75	57
Une	troupe	théâtrale	CCVC	113	2	fréquent	pt	388	72	1
Une	jupe	satinée	CVC	22	1	fréquent	ps	396	75	57
Une	coupe	chinoise	CVC	25	19	rare	pʃ	373	6	1
Une	soupe	concentrée	CVC	25	16	rare	pk	388	21	1
Une	lampe	chauffante	CVC	72	9	rare	pʃ	364	6	1
Une	nappe	charmante	CVC	22	9	rare	pʃ	379	6	1
Un	type	chanceux	CVC	136	8	rare	pʃ	337	6	1
Une	croupe	couverte	CCVC	9	7	rare	pk	388	21	1
Une	trompe	coupée	CCVC	8	7	rare	pk	425	21	1
Une	grippe	changeante	CCVC	6	5	rare	pʃ	436	6	1
Un	groupe	contrôle	CCVC	140	4	rare	pk	350	21	1
Une	grappe	chargée	CCVC	11	3	rare	pʃ	426	6	1
Une	troupe	canadienne	CCVC	113	2	rare	pk	374	21	1
Une	jupe	chiffonnée	CVC	22	1	rare	pʃ	393	6	1

Annexe 20 : Items de remplissage utilisés dans l'expérience 1.

blaʒid	diban	gœtɛb	sypad
bœlyv	drigun	ʒakɑ̃b	firʒal
bligyv	dɥilɔd	ʒibɑ̃v	gœrtɛb
ʃylit	flegiʒ	kratun	ʒalkɑ̃b
dakid	fiʒal	kobyp	ʒirbɔd

kɛptun	nozbɛd	zatkɛf	sartun
kiptaʒ	nyrlib	zaftim	
koʒbyp	pyʒban	zwatkub	
nifgɔʒ	tralbɔd	kytvin	

Annexe 21 : Items de remplissage utilisés dans les expériences 3 et 6.

ʒalkãb	sildum	sartun	tralbɔd
silgyn	fiʒal	dronim	driglyv
kratun	dɥilɔd	zaftim	
firʒal	dibraz	sypad	
sildaʒ	fikrav	magryl	
ʒakãb	dakid	drigun	

Annexe 22 : Items de remplissage utilisés dans l'expérience 4.

dakid	fiʒal	kratun	silgyn
diban	fikrav	krilan	sypad
dibraz	firʒal	magryl	tralbɔd
driglyv	glodraf	mirvad	zaftim
drigun	bladrin	sartun	ʃaktyp
dronim	ʒakãb	siglan	
dɥilɔd	ʒalkãb	sildaʒ	
bølyv	kiptaʒ	sildum	

Annexe 23 : Items de remplissage utilisés dans l'expérience 5.

ʒalkãb	sartun	diban	pyʒban
silgyn	dronim	siglan	tryvim
kratun	saftim	ʃaktyp	gørtɛb
firʒal	sypad	bladrin	kɔʒbyp
sildaʒ	magryl	krilan	kobyp
ʒakãb	drigun	pakliv	fyrsiʒ
sildum	tralbɔd	glibraʒ	vyldar
fiʒal	driglyv	zatkɛf	nyrlib
dɥilɔd	glodraf	bøgryv	ʃɔrdin
dibraz	kiptaʒ	gryviʒ	magdin
fikrav	bølyv	nɔzbɛd	blidryv
dakid	mirvad	kritryv	ʃygrid

dakriv
murkiv
fadrin
bligyv
kytvin

kɛptun
gøtɛb
zatkub
ʒibɑ̃v
blaʒid

kladyf
blugrid
nifkɔʒ
flegiʒ
sylpid

ʃylit
ʒirbɔd

Annexe 24 : Items de remplissage (*fillers*, à gauche) et 'attrapes' (*foils*, à droite) utilisés dans l'expérience 7.

Déterminant	Nom Commun	Adjectif
Une	robe	longue
Une	bague	dorée
Une	date	raisonnable
Une	botte	sale
Une	tente	robuste
Une	note	falsifiée
Un	stade	rayonnant
Une	ride	visible
Une	jambe	raide
Un	guide	rassurant

Déterminant	Nom Commun	Adjectif
Une	carte	pliée
Un	comte	ruiné
Une	cure	reposante
Un	coeur	fatigué
Une	cuisse	massive
Une	caste	détestée
Un	cône	lumineux
Une	cale	instable
Un	crime	horrible
Un	cube	radieux

Annexe 25 : Items de remplissage (*fillers*, à gauche) et 'attrapes' (*foils*, à droite) utilisés dans l'expérience 8.

Déterminant	Nom Commun	Adjectif
Une	grotte	récente
Un	geste	fautif
Un	globe	lumineux
Une	joute	soutenue
Une	fougue	rarissime
Une	meute	solidaire
Une	flûte	chantante
Une	figue	délicieuse
Un	monde	révoltant
Une	ronde	joyeuse
Une	veste	réversible
Une	ligue	dominante
Une	mouette	rieuse
Une	drogue	dangereuse
Une	fugue	laborieuse
Un	buste	mordoré

Déterminant	Nom Commun	Adjectif
Une	partie	difficile
Un	prêtre	silencieux
Une	palme	mouillée
Une	page	jaunie
Une	pomme	succulente
Une	poule	endormie
Un	parc	aéré
Un	prisme	lumineux
Une	porte	coincée
Une	peste	ravageuse
Une	pente	glissante
Une	piste	effacée
Une	pile	rechargée
Une	planche	fragile
Une	pêche	miraculeuse
Un	plateau	élevé

Résumé

Selon des travaux récents, les locuteurs seraient capables de recourir à des connaissances sur les régularités phonotactiques ou syllabiques de leur langue pour segmenter le signal de parole en mots. Une analyse distributionnelle de la fréquence des groupes de consonnes dans un lexique informatisé de la langue française révèle cependant l'existence d'un lien étroit entre la structure phonologique des groupes de consonnes et leur probabilité d'occurrence dans la langue. Les séquences de phonèmes tautosyllabiques ou légales sont en moyenne beaucoup plus fréquentes que les séquences hétérosyllabiques ou illégales. Trois modèles des processus de segmentation lexicale sont présentés qui pourraient rendre compte des effets observés sans recourir à des processus de segmentation fondés sur les régularités phonologiques de la langue. Les expériences conduites ont pour objectif de dissocier ces diverses interprétations d'un modèle de segmentation fondé sur les régularités phonologiques. Les résultats obtenus confirment en partie l'influence des contraintes phonologiques sur les processus de segmentation de la parole en mots. Ils incitent cependant à développer une réflexion sur le niveau d'implémentation de ces connaissances (prélexical vs. post-lexical) et sur leur mode d'intégration dans un modèle du traitement de la parole. Un modèle est proposé dans lequel ces procédures de segmentation constitueraient des routines générales d'organisation perceptive qui ne seraient pas spécifiques des processus de traitement du langage parlé et n'influenceraient pas directement les niveaux de représentation lexicaux. Ces deux catégories de traitements seraient indépendantes et intégrées au cours d'une étape ultérieure.

Mots-Clés

Perception de la parole, Segmentation lexicale, Régularités phonologiques, Régularités probabilistes, Compétitions lexicales

Title (translated)

Phonotactic regularities and the segmentation of speech into words:
Phonological, probabilistic or lexical effects?

Summary

It has been argued that speakers would be able to use their knowledge about phonotactic or syllabic regularities in the process of segmenting speech into words. However, a distributional analysis of the frequency of consonant clusters in a French computerised lexicon reveals a strong relationship between the phonological structure of consonant sequences and their frequency of occurrence in the language's words. On the average, tautosyllabic or phonotactically legal sequences are much more frequent than heterosyllabic or phonotactically illegal sequences. Three lexical segmentation models are discussed that could account for these effects without involving the influence of phonological regularities in speech segmentation processes. Experiments were conducted in order to test between the hypothesis of an involvement of phonological knowledge into lexical segmentation processes and these three alternate models. The data partly confirm the role of phonological constraints in the segmentation of speech into words. Nevertheless, they provide the basis to discuss issues of implementation level (pre-lexical vs. post-lexical) as well as the way prelexical and lexical informations should be integrated together. It is argued that these segmentation procedures would be performed according to general perceptual organisation principles that are not to be considered as language-specific. Indeed, prelexical segmentation processes would not influence lexical representations; both would be integrated at a later stage of processing.

Key-Words

Speech perception, Lexical segmentation, Phonological regularities, Probabilistic regularities, Lexical competition

Discipline

Psychologie

Unité de rattachement

Laboratoire de Psychologie Expérimentale (UMR 8581)
CNRS - Université Paris V / René Descartes
Centre Universitaire de Boulogne
71 Avenue Edouard Vaillant
92774 Boulogne-Billancourt Cedex

http://lpe.psycho.univ-paris5.fr