Université libre de Bruxelles
Faculté des Sciences Psychologiques et de l'Education
Laboratoire de Psychologie Expérimentale

Rôle des indices acoustico-phonétiques dans la segmentation lexicale: Etudes sur le français

Thèse présentée en vue de l'obtention du grade de Docteur en Sciences Psychologiques et de l'Education par

Nicolas Dumay

Promoteurs: Alain Content et Monique Radeau

Jury composé de:
Didier Demolin, Uli H. Frauenfelder, Régine Kolinsky,
José Junça de Morais et Juan Seguí

Année académique 2005-2006

Remerciements

Cette recherche a été financée par un mandat d'aspirant attribué à l'auteur par le Fonds National de la Recherche Scientifique belge. Je remercie Laetitia Desmet et Cédric Hanon pour leurs voix, Michel Dubray pour son assistance lors des enregistrements, et Mireille Cluytens pour avoir re-mesuré après moi les durées segmentales des stimuli de l'Expérience 4.

Mais avant tout, je tiens à remercier Alain Content pour avoir suivi ce travail d'un bout à l'autre, et Monique Radeau pour ses suggestions et ses encouragements.

Je remercie également Uli Frauenfelder pour m'avoir accueilli dans son laboratoire pour un stage pré-doctoral à l'Université de Genève.

Merci aussi à Didier Demolin, Marc Dominicy, Cécile Fougeron, Uli Frauenfelder et Régine Kolinsky pour des discussions essentielles, et à Monique Declercq, Michel Dumay, Fabienne Martin et Carolina Schultz pour avoir relu les versions antérieures de ce manuscrit.

Je tiens également à remercier Gerry Altmann, Paul Bertelson, Maud Boyer, Céline Chéreau, Alexandra Cleland, Monique Declercq, Adeline Dumay, Maxime Dumay, Michel Dumay, Eliane Gotteaux, Uli Frauenfelder, Graham Hitch, Marielle Lange, Julia Rayner, Carolina Schultz, Melanie Smeaton et Meesha Warmington pour leur soutien et leurs encouragements, ainsi que Gareth Gaskell et Markus Damian, pour m'avoir fait confiance et engagé comme chercheur sur les fonds du MRC et du BBSRC (Royaume Uni).

Enfin, je voudrais remercier Didier Demolin, Uli Frauenfelder, Régine Kolinsky, José Junça de Morais et Juan Seguí pour avoir accepté de faire partie du jury.

Bristol, Juin 2006

A mes parents

Table des matières

Avant-propos 9

Chapitre 1 Architecture du système de traitement de la parole et segmentation lexicale 11

1 Décoder le signal 11
1.1 Organisation perceptive de la parole 13
1.2 Elaboration d'une représentation d'entrée 14
2 Reconnaître les mots 17
2.1 Activation multiple et compétition lexicale 18
2.2 Nature de l'information sensorielle modulant l'activation lexicale 20
2.3 Nature des représentations lexicales 23
2.4 Flux d'information dans le système de reconnaissance des mots parlés 26
2.4.1 Effets du contexte lexical sur la perception? 27
2.4.2 Effets du contexte phrastique sur l'identification des mots? 29
3 Interpréter l'énoncé et construire un modèle du discours 30

Chapitre 2 Les mécanismes de la segmentation lexicale: théories et éléments empiriques 33

1 Approche lexicale: la segmentation comme produit de la reconnaissance des mots ... 33
1.1 La reconnaissance séquentielle des mots 33
1.2 Le processus de compétition lexicale 36
1.2.1 Effets de compétition lexicale 39
1.2.2 Limites de la compétition lexicale en tant que solution à la segmentation .. 42
2 Approche prélexicale: l'exploitation d'indices de segmentation 43
2.1 Indices acoustico-phonétiques 43
2.1.1 Assignation d'un phonème au mot qui précède ou à celui qui suit 44
2.1.1.1 Aperçu des variations phonétiques 44
2.1.1.2 Etudes perceptives basées sur les choix de segmentation explicites 46
2.1.1.3 Etudes perceptives utilisant une tâche de segmentation *online* 49
2.1.2 Attribution de deux syllabes au même mot ou à des mots différents 51
2.1.2.1 Aperçu des variations phonétiques 51
2.1.2.2 Etudes perceptives basées sur les choix de segmentation explicites 51
2.1.2.3 Etudes perceptives utilisant une tâche de segmentation *online* 53
2.1.3 En synthèse 56
2.2 Indices syllabiques 58
2.3 Indices phonotactiques et probabilités transitionnelles 60
2.3.1 Restrictions phonotactiques 60
2.3.2 Phonotactique probabiliste 61
2.3.3 Probabilités transitionnelles non-phonémiques 63
2.4 Indices rythmiques 64
2.4.1 Qualité vocalique (en anglais) 64

2.4.2 Accent lexical (en néerlandais) 65
2.4.3 Accent métrique (en français) 66
3 Approches combinées et hypothèse de "Possible Word Constraint" 67

Chapitre 3 Esquisse de la contribution expérimentale 70

1 Arrêt sur image 70
2 Considérations méthodologiques 73
3 Survol de la contribution expérimentale 75

Chapitre 4 Indices acoustico-phonétiques des frontières de mots: expériences sur les groupes obli 78

Expérience 1 – Effets d'alignement initial et final en word spotting 79
Méthode 79
Résultats et discussion 82
Expérience 2 – Perception des frontières de mots en tâche de catégorisation 89
Méthode 89
Résultats et discussion 91
Expérience 3 – Incidence des frontières de mots sur la syllabation explicite: données obtenues en tâche de répétition partielle 93
Méthode 94
Résultats et discussion 96
Analyse corrélationnelle entre word spotting, catégorisation et répétition partielle 100
Discussion générale 103

Chapitre 5 Nature du groupe consonantique et indices acoustico-phonétiques des frontières de mots: comparaison entre les groupes obli et "/s/ + OBstruante" 108

Expérience 4 – Effet d'alignement final en word spotting 109
Méthode 109
Résultats et discussion 110
Expérience 5 – Perception des frontières de mots en tâche d'appariement co-déterminé 118
Méthode 119
Résultats et discussion 120
Expérience 6 – Production d'énoncés OBLI et SOB isolément et en contexte de phrase 124
Méthode 124

Résultats et discussion 126
Expérience 6b – Comparaison entre frontières OBLI et frontières OBOB produites en contexte de phrases 131
Méthode 131
Résultats et discussion 132
Discussion générale 133

Chapitre 6 Compétition lexicale et indices acoustico-phonétiques des frontières de mots 136
Expérience 7 – Effet d'alignement final et compétiteur chevauchant 137
Méthode 137
Résultats et discussion 142
Expérience 8 – Effet d'alignement final, compétiteur enchâssant, et compétiteur du compétiteur 145
Méthode 145
Résultats 145
Analyses jointes des Expériences 7 et 8 147
Discussion générale 148

Chapitre 7 Conclusion générale 151

Bibliographie 155

Table des tableaux

Tableau 1. Temps de réaction moyens (TRs) mesurés à partir de la fin de la cible, et taux d'erreurs moyens (Ers) pour l'Expérience 1. Les écarts-types figurent entre parenthèses. 83
Tableau 2. Durées segmentales moyennes (ms) en fonction de la structure phonologique du porteur et de la position de la frontière lexicale (#). Les écarts-types figurent entre parenthèses. 84
Tableau 3. Répartition des réponses en fonction de l'assignation syllabique des consonnes du groupe médian. 96
Tableau 4. Proportion moyenne de réponses .CC (et écarts-types) en fonction de la partie à répéter, de la structure phonologique des stimuli et de la position de la frontière lexicale. 97
Tableau 5. Corrélations (r de Bravais-Pearson) entre l'effet d'alignement sur les TRs en *word spotting*, le taux de réponses correctes en catégorisation, et l'effet de la position de la frontière lexicale en répétition de première et de seconde partie, séparément pour les items VCCVC et CVCCV. 101
Tableau 6. Temps de réaction moyens (TRs) mesurés à partir de la fin de la cible, et taux d'erreurs moyens (Ers) pour l'Expérience 4. Les écarts-types figurent entre parenthèses. 111
Tableau 7. En haut: Durées segmentales moyennes (ms) en fonction de la nature du groupe consonantique et de la position de la frontière lexicale (#), telles qu'évaluées par l'auteur (N.D.) et par le juge indépendant (M.C.). En bas: Valeurs du coefficient de corrélation (r de Bravais-Pearson) entre les durées segmentales rapportées par l'auteur et celles rapportées par le juge indépendant. 114
Tableau 8. Durées segmentales moyennes (ms) en fonction du contexte de réalisation, du groupe consonantique et de la position de la frontière lexicale (#). Les écarts-types figurent entre parenthèses. 126
Tableau 9. Durées segmentales moyennes (ms) en fonction de la locutrice et de la nature du groupe consonantique. Les écarts-types figurent entre parenthèses. 132
Tableau 10. Durées segmentales moyennes (ms) en fonction de la position de la frontière lexicale (#) et du statut lexical de la portion finale CCV(C). Les écarts-types figurent entre parenthèses. 139
Tableau 11. Temps de réaction moyens (TRs) mesurés à partir de la fin de la cible, et taux d'erreurs moyens (Ers) pour l'Expérience 7. Les écarts-types figurent entre parenthèses. 142
Tableau 12. Temps de réaction moyens (TRs) mesurés à partir de la fin de la cible, et taux d'erreurs moyens (Ers) pour l'Expérience 8. Les écarts-types figurent entre parenthèses. 146
Tableau 13. En haut: Valeurs attendues du temps de réaction pour onze délais s'échelonnant de 0 à 800 ms. En bas: Temps de réactions observés, moyenne et écart-type (ms) sur 20 essais/valeur attendue, obtenus avec le dispositif des Expériences 3, 7 et 8. 251

Table des figures

Figure 1. Une architecture du système de traitement de la parole. Les flèches en trait plein représentent les flux d'information bas-haut; celles en pointillé représentent les flux d'information haut-bas. 12
Figure 2. Niveaux d'activation des mots inclus dans la liste des candidats lexicaux produite par Shortlist au cours de la présentation de la séquence "ship inquiry". Dans cette simulation, le nombre maximal de candidats est fixé à 30 (reproduit à partir de Norris, 1994, Figure 8, avec la permission de l'auteur). 39
Figure 3. Corrélation (et droite de régression) entre l'effet d'alignement initial (e.g., ic#roche vs. i#croche) sur les latences de *word spotting* et la différence de durée de la liquide entre les conditions C#C et #CC. 87
Figure 4. Taux moyen de réponses correctes (et intervalle de confiance) en fonction de la structure phonologique des stimuli et de la position de la frontière lexicale pour l'Expérience 2. 91
Figure 5. Temps de réaction moyens en fonction de l'alignement et du groupe consonantique pour les deux catégories de participants (rapides vs. lents). 112
Figure 6. Taux (ou taux moyen) de réponses correctes (et intervalle de confiance) en fonction du groupe consonantique pour N.D. et les 24 participants naïfs. 120
Figure 7. Distributions des différences de durée entre la condition C#C et la condition #CC pour V_1 (en haut) et C_3 (en bas). 122
Figure 8. Différences de durée (ms) entre la condition #CC et la condition C#C, en fonction du contexte de réalisation et du groupe consonantique. Les différences significatives sont marquées d'un indice allant de 1 à 3 selon que le niveau de signification est inférieur à .05, .01, ou .001. L'indice de gauche est relatif à l'analyse par participant, celui de droite, à celle par item. 127
Figure 9. Distributions des TRs en fonction de l'alignement dans l'Expérience 1. A gauche: Cibles CVC enchâssées en finale. A droite: Cibles CVC enchâssées à l'initiale. 234
Figure 10. Distributions des TRs en fonction de l'alignement dans l'Expérience 4 (cibles CVC ou CCVC). A gauche: Porteurs à groupe consonantique OBLI. A droite: Porteurs à groupe consonantique SOB. 234
Figure 11. Distributions des TRs en fonction de l'alignement dans l'Expérience 7 (cibles CVC ou CCVC). A gauche: Porteurs à finale mot. A droite: Porteurs à finale pseudomot. ... 234
Figure 12. Distributions des TRs en fonction de l'alignement dans l'Expérience 8 (cibles CV ou CCV). A gauche: Porteurs à finale mot. A droite: Porteurs à finale pseudomot. 234

Avant-propos

Si l'on en croit notre expérience subjective du langage parlé, les mots se présentent à notre tympan comme des entités physiques séparées. Pourtant, l'analyse objective du signal acoustique révèle que la parole est le plus souvent continue. En raison de la coarticulation des segments phonétiques (e.g., Menzerath & de Lacerda, 1933; voir Farnetani, 1997, pour une vue d'ensemble), les mots s'enchaînent les uns aux autres et les propriétés acoustiques de leurs phonèmes chevauchent dans le temps celles des phonèmes adjacents (e.g., Fant, 1962; Liberman, 1970; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). En outre, il ne semble y avoir, entre les mots, aucun élément acoustique systématique qui servirait à signaler les frontières (e.g., Cole & Jakimik, 1980a; Lehiste, 1972; Trubetzkoy, 1939). Par ailleurs, le flux de parole ne véhicule aucune information sémantique, et seule la reconnaissance des éléments constituant l'énoncé —c'est-à-dire les mots, auxquels nous pouvons assigner un sens— nous permet de déduire la signification particulière du message (voir Jusczyk & Luce, 2002, pour une vue d'ensemble). Comprendre un énoncé implique donc, pour l'auditeur, de le segmenter en ses constituants lexicaux, ou, en d'autres termes, de localiser dans l'onde acoustique l'endroit où les mots commencent et celui où ils se terminent. Plus précisément, il s'agit pour l'auditeur de trouver l'alignement correct entre les unités signifiantes contenues dans le signal, et ses propres représentations lexicales stockées en mémoire à long-terme. Par "alignement", nous voulons dire que l'auditeur doit savoir à quelle partie de représentation lexicale comparer chaque portion de l'entrée sensorielle. Le but du présent travail est d'explorer de manière analytique certains des mécanismes par lesquels l'humain confronte ainsi sa mémoire des mots à la nature continue de la parole et segmente —la plupart du temps adéquatement— les énoncés.

L'introduction à cette thèse comprend trois chapitres et se structure de la manière suivante. Dans le chapitre 1, nous tentons de donner une vue d'ensemble de l'architecture fonctionnelle du système de perception et de traitement des mots parlés. Après avoir isolé les différents *loci*

où opéreraient les mécanismes de segmentation lexicale[1], nous revoyons brièvement les étapes majeures menant à la reconnaissance des mots et à la compréhension du message. Dans le chapitre 2, la problématique de la segmentation est approfondie. Nous présentons les mécanismes par lesquels l'auditeur pourrait segmenter la parole, en tenant compte du niveau de traitement auquel ces mécanismes interviennent. Enfin, au chapitre 3, nous confrontons deux hypothèses mutuellement exclusives, et donnons un aperçu de la contribution expérimentale qui cherchera à les départager.

[1] L'emploi du terme "segmentation lexicale" n'implique pas que les mécanismes mis en jeu dans la segmentation de la parole soient nécessairement dépendants du lexique; ce terme signifie simplement "segmentation en mots", par opposition à segmentation en unités infra-lexicales, c'est-à-dire plus petites que le mot. Parfois, pour alléger le texte, le qualificatif "lexical" sera volontairement omis.

Chapitre 1 Architecture du système de traitement de la parole et segmentation lexicale

Au contraire des autres fonctions majeures que l'on peut assigner au système de traitement de la parole (décodage du signal, reconnaissance des mots, interprétation de l'énoncé, et intégration dans un modèle du discours), la segmentation lexicale ne correspond pas à une étape de traitement clairement distincte à l'intérieur du système. Grâce à certains mécanismes exclusivement dédiés à la segmentation, l'auditeur exploite de nombreux aspects de la chaîne parlée pour localiser les frontières de mots. Cependant, la manière dont il perçoit au final comment se partitionnent les énoncés dépend aussi, dans une large mesure, des opérations mises en jeu lors de la reconnaissance des mots et de l'interprétation du message. Dans la Figure 1, qui schématise l'architecture cognitive de ce système, la contribution à la segmentation lexicale des opérations impliquées dans la reconnaissance des mots, tout comme dans l'analyse syntaxico-thématique, est représentée par l'ancrage de ces dernières dans la zone délimitée par les deux traits discontinus. En amont du traitement (voir le bas du schéma), la mise en branle de mécanismes de segmentation spécifiques et pré-lexicaux, basés sur l'occurrence dans le signal d'éléments corrélés avec les frontières de mots —que l'on appelle "indices de segmentation"— ne peut avoir lieu tant que le décodage perceptif de ces éléments n'est pas terminé. En revanche, en aval du traitement (voir le haut du schéma), le déclenchement des mécanismes de reconnaissance des mots ne requiert pas que les frontières lexicales aient été préalablement localisées. Pour mieux cerner la fonction de segmentation et comprendre la mécanique générale du système dans lequel elle s'inscrit, nous donnons un aperçu des autres composantes du système de traitement de la parole (en allant du bas vers le haut).

1 Décoder le signal

Le décodage du signal est la première étape du traitement et consiste essentiellement en deux opérations: organiser perceptivement la parole et élaborer une représentation d'entrée capable d'activer notre mémoire à long-terme pour les mots (ou lexique mental).

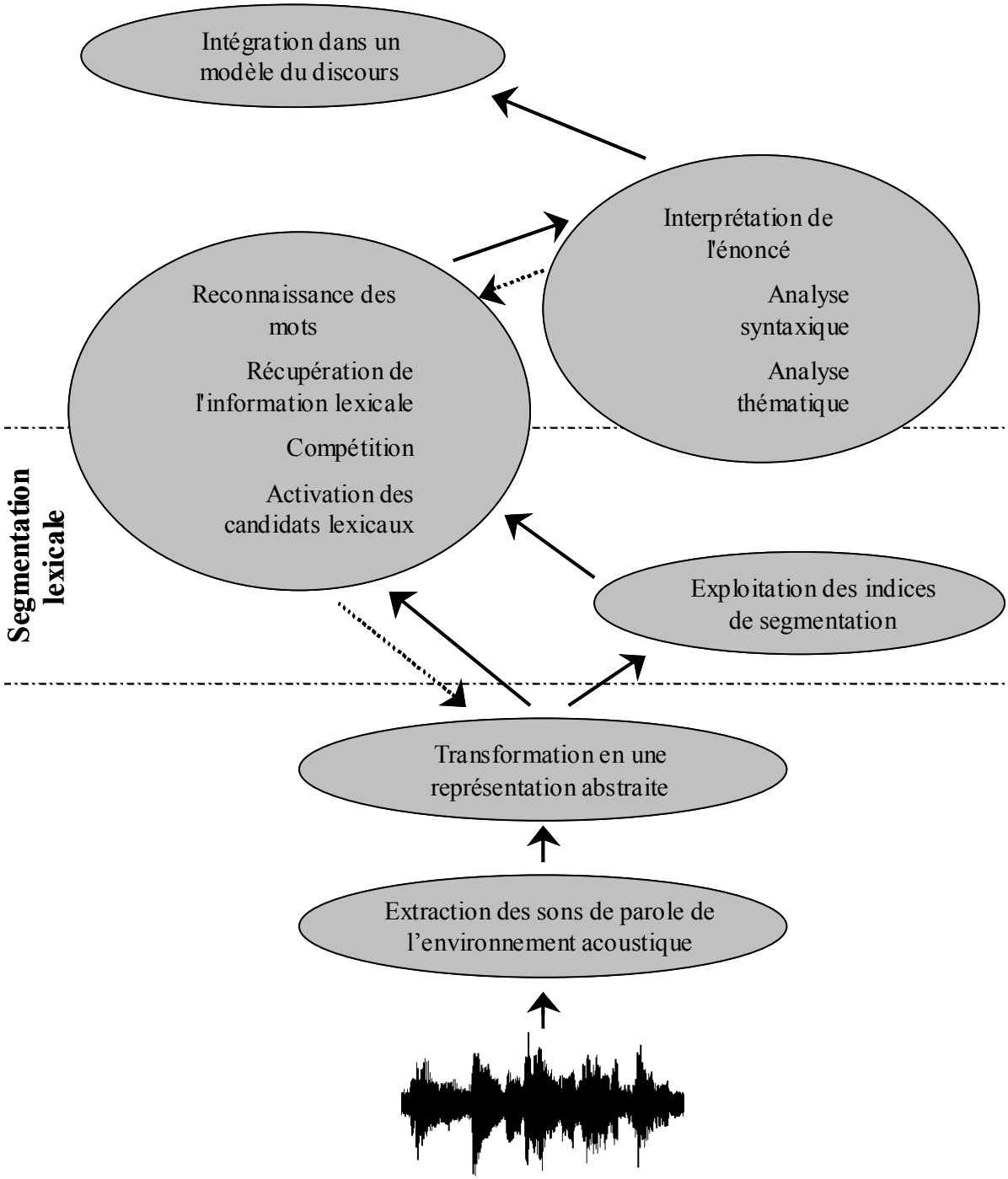

Figure 1. Une architecture du système de traitement de la parole. Les flèches en trait plein représentent les flux d'information bas-haut; celles en pointillé représentent les flux d'information haut-bas.

1.1 Organisation perceptive de la parole

Les sons de parole de notre interlocuteur ne parviennent pas à notre oreille par un chenal qui leur serait exclusif. Dans le signal acoustique qui les véhicule jusqu'à nous s'y entremêlent tous les autres sons présents au même moment dans l'environnement. La première opération impliquée dans le décodage consiste donc à extraire de cet imbroglio acoustique les éléments qui correspondent à de la parole et à les assigner au locuteur qui les a produits (voir Cooke & Ellis, 2001, pour un état de la question). Dans le domaine, la théorie la plus complète est celle de Bregman (1990), pour qui les mécanismes d'organisation perceptive de la parole ne seraient pas différents de ceux mis en jeu lors de la perception de tout autre événement sonore (voir Darwin, 1997, pour un aperçu). Dans un premier temps, les diverses composantes fréquentielles du spectre neural seraient réparties en objets auditifs cohérents, grâce à l'intervention de processus analytiques innés, obligatoires, rapides et strictement perceptifs[2], obéissant aux principes de la Gestalt (de similarité, de proximité, de clôture, de co-modulation et de continuité; cf. Wertheimer, 1923) et appliqués aux trois dimensions suivantes: fréquence, amplitude et localisation. Dans un second temps, des processus d'intégration basés sur des schémas acquis, et dont certains pourraient être spécifiques à la parole (comme par exemple, une connaissance des caractéristiques du tractus vocal et des contraintes qu'elles imposent à la production des sons) viendraient corriger, le cas échéant, l'attribution erronée d'éléments dissimilaires à différents percepts auditifs.

Cependant, les hypothèses de Bregman (1990) ne semblent pas faire l'unanimité, du moins pour la parole. D'une part, il n'y a pas d'argument solide en faveur de l'existence de processus tardifs d'organisation perceptive impliquant la reconnaissance de schémas. D'autre part, de nombreux phénomènes ne respectant pas les principes de ségrégation primaires dérivés de la Gestalt ont été mis en évidence (Remez, Rubin, Berns, Pardo, & Lang, 1994). Le fait que dans ces derniers, le percept final semble lié à l'occurrence de variations sensorielles complexes au cours du temps, comme, par exemple, celles de l'enveloppe spectrale (Remez, Yang, Piorkowski, Wissig, Batchelder, & Nam, 2002; Smith, Delgutte, & Oxenham, 2002), plutôt

[2] Par "processus […] strictement perceptifs", nous entendons des processus exclusivement guidés par l'entrée sensorielle.

qu'au caractère naturel et à la qualité acoustique de chaque composante a d'ailleurs conduit certains auteurs à postuler que la parole devait faire l'objet de mécanismes d'organisation perceptive tout à fait spécifiques (voir Remez, 2003; Rosenblum, 2005). Selon Remez (2003), les démonstrations récentes montrant que l'intégration phonétique intermodale peut avoir lieu en dépit d'une incongruence temporelle (Munhall, Gribble, Sacco, & Ward, 1999) ou spatiale (Bertelson, Vroomen, & de Gelder, 1997) substantielle entre l'auditif et le visuel, pourraient même suggérer que les mécanismes en question soient de nature supramodale. La problématique de l'organisation perceptive de la parole est donc loin d'être résolue.

1.2 Elaboration d'une représentation d'entrée

Une fois que les éléments du spectre qui se rapportent à de la parole ont été isolés et regroupés en un continuum sensoriel cohérent (i.e., celui produit par le locuteur), la seconde opération du décodage peut commencer. Celle-ci est incontestablement linguistique, et consiste en l'élaboration d'une représentation pré-lexicale de l'entrée acoustique servant notamment de code d'accès et d'appariement au lexique mental. Puisque la structure de la représentation d'entrée doit être isomorphe à celle de toute entrée lexicale pour que l'appariement de l'une à l'autre soit possible, la manière de concevoir la première est nécessairement liée à l'idée que l'on se fait de la seconde, et vice-versa.

Pour la plupart des théories de la reconnaissance des mots parlés, les entrées lexicales sont des entités relativement abstraites, qu'elles soient chacune la représentation canonique d'un mot (Gaskell & Marslen-Wilson, 1997; Lahiri & Marslen-Wilson, 1991; McClelland & Elman, 1986; Norris, 1994; Pisoni & Luce, 1987; Stevens, 2002; Studdert-Kennedy, 1976), ou l'une de ses variantes phonologiques, lesquelles seraient toutes représentées (Harrington, Laver, & Cutting, 1986; Klatt, 1979). L'hypothèse d'un lexique abstrait est motivée par l'extrême variabilité du signal. En effet, la forme acoustique des segments est affectée par de nombreux facteurs, parmi lesquels le contexte phonétique, du fait de la coarticulation (voir plus haut; Liberman, Delattre, Cooper, & Gertsman, 1954), le taux d'articulation (Port, 1981), la prosodie de la phrase, sans oublier les caractéristiques du locuteur, comme, par exemple, la taille de son tractus vocal (Peterson & Barney, 1952), sa fréquence glottale (Monsen &

Engebretson, 1977), ou encore son accent dialectal. En outre, jusqu'à présent, on a trouvé peu d'invariants acoustiques qui permettent d'identifier correctement les segments phonétiques (néanmoins, voir Cole & Scott, 1974; Stevens, 2002; Stevens & Blumstein, 1981). Dès lors, l'élaboration de la représentation d'entrée implique généralement un processus d'abstraction et de normalisation, appelé "classification", au travers duquel les portions successives du signal sont transformées en unités infra-lexicales épurées des détails acoustiques non-pertinents. Ce processus est basé en partie sur un principe d'économie: au lieu d'essayer d'apparier directement chaque portion singulière de l'entrée sensorielle à une des multiples entrées lexicales, il est plus avantageux de recoder d'abord l'entrée acoustique en un nombre fini d'unités élémentaires plus petites. Néanmoins, ceci ne signifie pas que le processus de classification ait pour objectif de produire une représentation d'entrée qui équivaudrait à la simple concaténation d'unités phonologiques monolithiques. La parole comporte de nombreuses variations qui, d'un point de vue linguistique, ont une valeur informative prépondérante (cf. la section 2.2). Par conséquent, la représentation d'entrée est vraisemblablement très détaillée, comportant plusieurs niveaux de description et se caractérisant par une structure interne complexe.

D'après Norris et Cutler (1985), il faut au moins que trois prérequis soient satisfaits pour qu'un fragment de parole puisse fonctionner comme unité de classification. Primo, les catégories du fragment doivent être suffisamment discriminables les unes des autres dans le signal. S'il n'est pas nécessaire que ces catégories soient plus discriminables que les mots eux-mêmes, il importe que l'ensemble de catégories envisagé soit considérablement plus petit que le nombre de mots dans le lexique. Secundo, il doit être possible de décrire n'importe quel énoncé comme une combinaison d'instances du fragment en question, sans qu'aucune portion du signal ne soit laissée pour compte. Par exemple, puisqu'il est impossible de décrire complètement un énoncé uniquement sur base des consonnes, ces dernières ne peuvent servir de code d'accès. Enfin, tertio, et plus pertinent pour la problématique de la segmentation lexicale, les frontières entre les unités de classification doivent correspondre —au moins pour certaines d'entre elles— aux frontières de mots. Le phonème est une unité qui remplit *de facto* cette condition, puisque les mots commencent et se terminent toujours à une frontière entre

deux phonèmes. En revanche, cette condition n'est pas satisfaite par une unité comme le diphone, c'est-à-dire la portion qui s'étend du milieu d'un segment au milieu du segment suivant et dont les frontières, par définition, ne coïncident jamais avec les frontières de mots. De ce point de vue, le calcul d'unités de classification pourrait simplifier le processus de segmentation. En effet, que les recherches lexicales soient initiées au début de chaque nouvelle unité de classification alors que l'identité de cette dernière n'est pas encore connue, ou que les candidats lexicaux compatibles ne soient activés qu'au terme de l'analyse prélexicale, dans un comme dans l'autre, le nombre d'appariements superflus peut être, en principe, fortement réduit. On voit donc que le moment où débute l'appariement, tout comme l'étendue du champ lexical activé, dépend de la taille de la représentation d'entrée.

La question du format de cette représentation est toujours l'objet de controverses. Le phonème étant la plus petite unité en termes de laquelle la forme phonologique de tout énoncé peut être décrite séquentiellement, cette unité linguistique a longtemps été considérée comme le seul format possible de la représentation d'entrée (cf. Chomsky & Halle, 1968; voir Manaster-Ramer (1996) et Port (1996), pour une discussion). Par la suite, les positions théoriques se sont diversifiées, et à côté du phonème (Foss & Gernsbacher, 1983; McClelland & Elman, 1986; Pisoni & Luce, 1987; Slowiaczek, Nusbaum, & Pisoni, 1987; Slowiaczek, McQueen, Soltano, & Lynch, 2000), ont été proposées d'autres unités, pour la plupart, elles aussi, de type linguistique, telles que les gabarits spectraux (Klatt, 1979), les traits phonétiques distinctifs —évalués de manière discrète (Lahiri & Marslen-Wilson, 1991; Stevens, 2002) ou continue (Marslen-Wilson & Warren, 1994)—, les allophones positionnels (Luce, Goldinger, Auer, & Vitevitch, 2000), les demi-syllabes (Fujimura & Lovins, 1978; Samuel, 1989), les diphones (Marcus, 1981), les syllabes (Church, 1987; Mehler, Dommergues, Frauenfelder, & Seguí, 1981; Mehler, Dupoux, & Seguí, 1990; Seguí, 1984; Seguí, Dupoux, & Mehler, 1990), ou encore les syllabes accentuées (Grosjean & Gee, 1987; Mattys & Samuel, 1997, 2000).

Certaines théories proposent toutefois une solution différente aux problèmes que génère le manque d'invariance dans le continuum acoustique. C'est le cas, notamment, de la Théorie Motrice de la parole, proposée par Liberman et collaborateurs. Suivant celle-ci, l'auditeur

serait à même de dériver le seul invariant véritable sous-jacent à tout phonème, à savoir le geste phonétique du locuteur. Alors que dans la version originale de cette théorie (Liberman *et al.*, 1967), les invariants correspondraient aux commandes motrices à la source des gestes articulatoires correspondant aux unités phonétiques, dans la version révisée (Liberman & Mattingly, 1985, 1989; Liberman & Whalen, 2000), les invariants résideraient dans les structures plus abstraites et intentionnelles qui contrôlent ces mouvements (voir Klatt, 1989, pour une discussion sur les limites de cette approche; voir aussi Fowler, 1986, 1990; Fowler & Rosenblum, 1991, pour une théorie connexe). Un autre exemple est fourni par la théorie de la reconnaissance des mots par analogie, mise en avant par Goldinger (1996, 1998, 2000; voir aussi Goldinger, Kleider, & Shelley, 1999). Selon cette théorie, les mots ne seraient pas stockés dans le lexique sous une forme prototypique, et donc abstraite, mais sous la forme des multiples traces épisodiques, très précises, résultant de toutes les situations dans lesquelles le mot a été entendu. Dans cette perspective, l'appariement lexical consisterait donc à trouver, parmi l'ensemble des traces acoustiques présentes en mémoire à long-terme, celle qui ressemble le plus à la représentation "analogique" du mot à l'entrée. Plus le nombre d'exemplaires activés par cette dernière est important —comme c'est par exemple le cas pour un mot fréquent—, plus l'écho (ou la réponse) émergeant du système lexical serait de nature abstraite. La question du format et de la texture de la représentation d'entrée reste donc ouverte.

2 Reconnaître les mots

La reconnaissance des entités lexicales présentes dans le signal acoustique est la clé du processus de compréhension du message, puisqu'elle permet à l'auditeur d'accéder aux propriétés morphologiques, sémantiques et syntaxiques des mots stockées en mémoire à long-terme (Oldfield, 1966). En appariant ainsi l'entrée sensorielle aux représentations de mots (ou entrées lexicales) les plus appropriées, l'auditeur est à même de rattacher un sens à une forme, et vice versa.

2.1 Activation multiple et compétition lexicale

Etant donné que les mots d'une langue sont construits sur base d'un répertoire de seulement 30 à 40 phonèmes en moyenne (Maddieson, 1984) et qu'ils sont donc fortement similaires, la reconnaissance des mots consiste essentiellement en un processus de discrimination entre représentations lexicales. De ce fait, la plupart des modèles de reconnaissance actuels supposent que les multiples candidats lexicaux compatibles avec une même portion de l'entrée acoustique sont activés (ou évoqués) simultanément et de manière automatique, et en compétition active les uns avec les autres pour la reconnaissance. L'hypothèse d'activation simultanée est présente dans tous les modèles, et ce depuis la version initiale de la théorie de la Cohorte (Marslen-Wilson & Welsh, 1978). Dans cette dernière, l'activation est représentée de manière binaire, les entités lexicales étant activées ou pas. En revanche, dans les autres modèles, l'activation varie de manière graduelle, en fonction du degré de compatibilité de la représentation lexicale avec l'entrée sensorielle, un postulat emprunté à la théorie (amodale) des Logogènes (Jackson & Morton, 1984; Morton, 1969, 1979).

L'hypothèse d'un mécanisme de compétition active entre les différents mots évoqués par le même fragment sensoriel a été mise en avant dans TRACE, modèle connexionniste à représentations locales hautement interactif (McClelland & Elman, 1986; voir aussi Frauenfelder & Peeters, 1990, 1998). Dans ce modèle, tout comme dans le modèle Shortlist (Norris, 1994) qui s'en inspire fortement sur ce point, le phénomène de compétition est implémenté au moyen du mécanisme d'inhibition latérale entre les compétiteurs lexicaux. Chaque unité lexicale affecte négativement le niveau d'activation de ses compétiteurs proportionnellement à son propre niveau d'activation et au poids de la connexion qui la lie au compétiteur. Le poids de cette connexion reflète notamment le degré de recouvrement temporel entre les deux unités. Dans NAM, i.e., *Neighborhood Activation Model* (Luce & Pisoni, 1998; Luce, Pisoni, & Goldinger, 1990; voir aussi Luce, 1986a) et dans PARSYN (Luce *et al.*, 2000), son instanciation connexionniste, la compétition lexicale prend une forme quelque peu différente, en ce qu'elle lie la notion de densité du voisinage de similarité à celle de fréquence d'occurrence des voisins (cf. Landauer & Streeter, 1973; Luce, 1989). La densité du voisinage d'un mot se réfère au nombre de mots phonétiquement proches et de même

longueur phonémique, parfois plus ou moins une unité. NAM postule que la probabilité d'identifier correctement un mot dépend de sa fréquence et, inversement, de la densité et de la fréquence de ses voisins. Dans ce modèle, la fréquence est conçue comme un biais décisionnel dont l'influence peut varier en fonction de la tâche (cf. Balota & Chumbley, 1984), et non comme une constante affectant le niveau de repos des représentations lexicales ou leur taux d'activation (e.g., Marlsen-Wilson, 1987). Enfin, dans la version actuelle de la théorie de la Cohorte (Gaskell & Marslen-Wilson, 1997, 1999, 2002), modélisée à l'aide d'un réseau récurrent simple (cf. Elman, 1990; Jordan, 1986; Norris, 1990), l'activité lexicale est représentée par le patron d'activation émanant d'un nombre fini (ou couche) d'unités codant les propriétés phonologiques du mot en cours d'évocation. A chaque instant, ce patron d'activation correspond au résultat de la fusion de l'ensemble des représentations lexicales compatibles avec l'entrée. Sous cet angle, le phénomène de compétition est donc totalement implicite, puisque l'interférence exercée par chaque candidat lexical sur les candidats évoqués simultanément est intrinsèque à la nature distribuée de la représentation.

Dans la théorie de la Cohorte, tout comme dans les modèles TRACE et Shortlist, la compétition lexicale est conçue comme un processus hautement directionnel, et ce parce que la parole est distribuée dans le temps: parmi l'ensemble des représentations lexicales qui rendent compte d'une portion de l'entrée sensorielle, seules celles qui correspondent à cette portion depuis leur segment initial acquièrent un niveau d'activation substantiel. De ce fait, seules ces représentations constituent les réels compétiteurs du mot qui a été articulé. A l'opposé, NAM n'accorde aucun statut particulier aux segments initiaux des représentations lexicales, ce modèle ayant été au départ élaboré pour rendre compte du traitement des mots courts. A fréquence d'occurrences équivalente, un mot correspondant à l'entrée sensorielle par ses segments non-initiaux affecte tout autant la reconnaissance qu'un mot similaire à l'entrée dès son premier segment.

Les hypothèses d'activation multiple et de compétition lexicale ont reçu de nombreuses confirmations empiriques. La question du degré de directionnalité dans le traitement reste cependant un élément de controverse. Comme nous le verrons au chapitre 2, section 1.2, la

compétition entre les représentations lexicales compatibles avec l'entrée sensorielle offre une solution générale et robuste au problème de la segmentation en mots.

2.2 Nature de l'information sensorielle modulant l'activation lexicale

Traditionnellement, on considérait que l'entrée sensorielle serait recodée en une chaîne de prototypes phonémiques ou en faisceaux constitués de leurs traits phonétiques sous-jacents (e.g., Stevens, 2002). Cette hypothèse est de plus en plus intenable, car de nombreux résultats indiquent au contraire que l'activation des candidats lexicaux est sensible à certains détails acoustiques subtils, de nature infra-phonémique (ou sous-catégorielle). La plupart des études se centrent sur le cas des voyelles, connues pour être perçues de façon moins catégorielle que les consonnes (Fry, Abramson, Eimas & Liberman, 1962), et sans doute traitées par des mécanismes perceptifs spécifiques (Schouten & van Hessen, 1992). Elles démontrent que la présence, en fin de segment vocalique, d'indices coarticulatoires incompatibles avec la consonne suivante perturbe l'accès lexical, et ce en décision lexicale et en catégorisation phonétique (Marslen-Wilson & Warren, 1994; Streeter & Nigro, 1979; Whalen, 1984, 1991, pour l'anglais; McQueen, Norris, & Cutler, 1999, pour le néerlandais), ou encore dans le paradigme d'analyse des mouvements oculaires (Dahan & Tanenhaus, 2004; Dahan, Magnuson, Tanenhaus, & Hogan, 2001). Par exemple, dans la dernière étude mentionnée, les mouvements oculaires des participants étaient examinés tandis que ceux-ci obéissaient à des instructions orales leur demandant de sélectionner un objet à l'écran à l'aide de la souris de l'ordinateur (e.g., "Now, click on the *net*."); par ailleurs, comme dans les autres études, l'incompatibilité portait sur le lieu d'articulation. Confirmant l'idée que l'information infra-segmentale est évaluée par le système de reconnaissance de manière continue dans le temps —mais aussi l'hypothèse de compétition lexicale—, Dahan *et al.* (2001) ont trouvé que les latences de fixation de l'objet-cible étaient plus longues lorsque le début du mot correspondant, y inclus sa voyelle, provenait d'un compétiteur lexical (e.g., "ne(ck)t") que lorsqu'il provenait d'un nonmot (e.g., "ne(p)t"), et ce, que l'image du compétiteur soit présente ou pas à l'écran (voir Dahan *et al.*, 2001; Gaskell & Marslen-Wilson, 1997; Norris, McQueen,

& Cutler, 2000; pour des simulations de ces effets sur TRACE, le modèle distribué de la Cohorte et Shortlist).

Dans le cas des consonnes, l'évidence empirique la plus pertinente est fournie par une étude récente de McMurray, Tanenhaus et Aslin (2002; voir aussi McMurray & Aslin, 2005, pour une étude connexe), manipulant la latence du voisement (ou VOT, *voice onset time*) en anglais. Dans cette langue, la latence du voisement est généralement définie comme l'intervalle de temps entre l'explosion consonantique et le premier cycle de vibrations laryngiennes. C'est pratiquement la seule dimension acoustique qui permette de distinguer la consonne voisée (dont le VOT est court) de la consonne non-voisée (dont le VOT est long) dans une paire de minimale de plosives (e.g., /g/ vs. /k/). McMurray *et al.* (2002) se sont inspirés des travaux d'Andruski, Blumstein et Burton (1994) montrant que lorsqu'une plosive à l'initiale d'un mot (e.g., "king") se caractérise par un VOT atypique (i.e., proche de la frontière catégorielle), ce mot n'évoque pas sa représentation lexicale, et, de ce fait, ne facilite pas la décision lexicale relative à un associé sémantique (e.g., "queen") autant que lorsque le VOT de son premier segment est prototypique (i.e., loin de la frontière catégorielle). Ayant pour but de déterminer si ces effets reflètent véritablement un gradient de sensibilité de la part du système de reconnaissance aux variations infra-phonémiques, McMurray *et al.* (2002) ont, eux aussi, exploité l'enregistrement des mouvements oculaires durant la tâche de sélection d'objets (cf. plus haut). Manipulant le VOT de la plosive initiale du stimulus auditif le long d'un continuum entre deux alternatives lexicales (e.g., "bear/pear"), ils ont trouvé que la probabilité de fixer l'image du compétiteur ("pear") aux dépens de la cible ("bear") augmente de manière linéaire à mesure que le VOT se rapproche de la frontière catégorielle. De petites différences au niveau du VOT n'ayant qu'un impact mineur sur l'identification phonémique modulent donc significativement, et par une fonction presque bijective, l'activation des candidats lexicaux. Dans la mesure où ces variations acoustico-phonétiques intra-catégorielles semblent corrélées avec l'environnement phonétique adjacent (e.g., Fougeron & Keating, 1997; Kessinger & Blumstein, 1998), il est raisonnable de penser qu'elles sont exploitées par le système de perception lors de la segmentation et de la reconnaissance des mots dans la parole continue. En relation directe avec ce point, les chapitres expérimentaux 4 et 5 auront

pour objectif de déterminer s'il existe des variations infra-segmentales liées à la position des frontières de mots en français, d'établir la généralité de ces indices, et d'explorer leur rôle dans la segmentation lexicale.

Dans certaines langues, l'activation lexicale dépend également du degré d'appariement avec l'information suprasegmentale (ou prosodique), c'est-à-dire avec les variations de la fréquence du fondamental (F0), de l'amplitude et/ou de la durée des constituants plus larges que le phonème, potentiellement pertinentes pour l'identité du mot (Cutler, Dahan, & van Donselaar, 1997, pour une première revue approfondie). Comme toute autre propriété phonologique, la structure suprasegmentale est exploitée si elle permet de réduire de manière significative et singulière l'incertitude du système de reconnaissance. Ainsi, l'information tonale —véhiculée uniquement par la F0— dans des langues telles que le mandarin, le cantonais ou le thaï, dont certains mots ne diffèrent que par la hauteur ou par le contour tonal de leurs syllabes, est partie prenante dans l'identification du mot (e.g., Burnham *et al.*, 1996; Ching, 1985, 1988; Cutler & Chen, 1995). En cantonais, par exemple, la séquence /si/ signifie "poème" au ton 1, "temps" au ton 2 et "histoire" au ton 6. Il en va de même du patron d'accentuation tonale en japonais (e.g., Cutler & Otake, 1996, 1999; Minematsu & Hirose, 1995; Otake, Yoneyama, Cutler, & van der Lugt, 1996), où l'on trouve de nombreuses paires de mots polysyllabiques ne différant que par la valeur haute ou basse du ton associé à chacune de leurs moras. Dans cette langue, la présentation d'un mot appartenant à une paire minimale par l'accent tonal (e.g., "amé" HB, *pluie*, vs. "amé" BH, *sucre*) ne facilite pas la reconnaissance subséquente de l'autre membre de la paire. Cela suggère qu'une différence de patron accentuel suffit à ce qu'un candidat lexical soit exclu de la compétition, voire ne soit pas activé du tout (Cutler & Otake, 1999). Cependant, l'information tonale, qu'il s'agisse du ton lexical ou du patron accentuel, semble traitée de manière plus lente et plus fragile que la chaîne segmentale, vraisemblablement parce que les variations pertinentes de F0 ne sont identifiées que lorsqu'une partie substantielle de la voyelle a été entendue (Cutler & Chen, 1997; Repp & Lin, 1990, Sekiguchi & Nakajima, 1999; Walsh Dickey, 1996). Dès lors, bien que ce type d'information puisse guider l'appariement lexical, sa contribution n'est peut-être pas

systématique, ou, en tout cas, pas aussi importante que celle de l'information segmentale et infra-segmentale.

De manière analogue, l'accent lexical (ou *stress*) oriente clairement l'appariement au lexique dans des langues telles que le néerlandais (Cutler & Koster, 2000; Cutler & van Donselaar, 2001; van Donselaar, Koster, & Cutler, 2005; Koster & Cutler, 1997; van Leyden & van Heuven, 1996) ou l'espagnol (Soto, Sebastián-Gallés, & Cutler, 2001). La question de savoir s'il exerce une fonction équivalente dans le cas de l'anglais est en revanche l'objet d'une controverse (cf. Mattys & Samuel, 2000). Dans cette langue, contrairement aux autres langues à accent lexical, la voyelle des syllabes inaccentuées est presque toujours réduite (en la voyelle schwa). En raison de cette corrélation, l'information segmentale permet donc presque tout le temps de rejeter un compétiteur lexical avant que ne le permette l'information strictement suprasegmentale, le contraste prosodique entre syllabes accentuées et inaccentuées n'ayant que peu d'impact sur l'activation des mots (Altmann, 1990; Altmann & Carter, 1989). De fait, en anglais, la plupart des études ayant manipulé l'occurrence de l'accent du mot indépendamment de la qualité vocalique n'ont pu mettre en évidence aucun rôle important de l'information suprasegmentale lors de l'appariement lexical (Cutler, 1986; Cutler & Clifton, 1984; Fear, Cutler & Butterfield, 1995; Slowiaczek, 1990, 1991). Toutefois, des résultats obtenus récemment dans des expériences variant la position relative des accents primaire et secondaire (dans des quadrisyllabes) vont à l'encontre de cette conclusion (Mattys, 2000; Mattys & Samuel, 1997, 2000).

En définitive, on voit donc que le système de traitement des mots exploite une représentation d'entrée très détaillée, incluant l'information segmentale et suprasegmentale théoriquement pertinente pour le processus de discrimination lexicale, mais aussi des indices acoustiques de bas niveau reflétant les modalités coarticulatoires des segments.

2.3 Nature des représentations lexicales

On peut concevoir les entités lexicales comme des associations mnésiques à long-terme de propriétés phonologiques (ou plus simplement acoustico-phonétiques), morphologiques, orthographiques, syntaxiques et sémantico-conceptuelles, à partir desquelles notre grammaire

mentale construit et reconstruit propositions et phrases (voir Jackendoff, 1997, pour une introduction). Si la plupart des entités lexicales sont de la taille du mot, rien n'empêche *a priori* que soient représentées des associations relatives à des items d'un domaine plus large, comme les idiomes, ou plus petit, comme les morphèmes. Bien entendu, le mode de représentation de ces informations peut influencer la manière dont on conçoit la reconnaissance des mots et la segmentation lexicale. Nous nous limiterons ici à évaluer cette question pour ce qui est des trois premières dimensions; le rôle des informations syntaxiques et sémantico-conceptuel sera quant à lui abordé dans la section 2.4.2.

En ce qui concerne la dimension phonologique, deux choses doivent être retenues. Primo, le manque d'invariance de la parole requiert un certain degré d'abstraction, comme nous l'avons vu à la section 1.2. L'abstraction est une caractéristique des *représentations*, si l'on postule que seules les formes prototypiques des mots ou leurs variantes contextuelles sont stockées (e.g., Pisoni & Luce, 1987). Elle est une propriété du *traitement,* et non des représentations, si l'on postule au contraire que notre mémoire des mots est exhaustive et contient la trace acoustique laissée par chaque événement lexical (cf., Goldinger, 1998). Secundo, le fait que l'activation lexicale soit sensible d'une manière linéaire à certaines variations infra-phonémiques (e.g., McMurray *et al.*, 2002) suggère que les représentations formelles sont aussi très détaillées.

En ce qui concerne la morphologie, les modèles qui traitent de ce type d'information (voir Schriefers, 1999, pour une introduction) envisagent la structure des entrées lexicales essentiellement de trois manières différentes. Selon certains modèles, la morphologie ne joue aucun rôle dans l'identification, et tous les mots, qu'ils soient morphologiquement simples ou complexes, possèdent une entrée lexicale (e.g., Butterworth, 1983; Grainger, Colé, & Seguí, 1991; Schriefers, Zwitserlood, & Roelofs, 1991). A l'opposé, d'autres modèles considèrent que le lexique d'entrée ne contient que les racines des mots et que, par conséquent, l'accès à un mot complexe requiert de le débarrasser au préalable de ses affixes (e.g., Marslen-Wilson, Ford, Older, & Zhou, 1996; Marslen-Wilson, Tyler, Waksler, & Older, 1994; Taft & Forster, 1975; Taft, Hambly & Kinoshita, 1986). Enfin, il existe aussi des modèles hybrides, à deux voies, qui supposent que tous les mots ainsi que chacun de leurs morphèmes ont une entrée

lexicale spécifique. Dans ces modèles, les mots complexes sont identifiés soit en passant par une étape pré-lexicale de décomposition affixes/racine, soit en accédant d'emblée aux formes complètes (e.g., Baayen, Dijkstra, & Schreuder, 1997; Caramazza, Laudanna, & Romani, 1988; Frauenfelder & Schreuder, 1992). Du point de vue de la segmentation, le domaine sur lequel se joue la compétition lexicale n'est donc pas le même selon qu'on emprunte la voie directe, où il correspond au mot, ou la voie impliquant une étape de décomposition en morphèmes, où il est de taille plus réduite. En outre, contrairement à la voie directe, l'étape de décomposition morphologique laisse entrevoir la possibilité que les frontières de morphèmes soient marquées d'une façon distincte dans le signal, et, donc, que les mots morphologiquement complexes soient extraits du continuum de parole par des mécanismes de segmentation spécifiques.

Enfin, en ce qui concerne l'orthographe, son influence sur le traitement des mots parlés a le plus souvent été démontrée au moyen de tâches métaphonologiques, comme le jugement de rime (Seidenberg & Tanenhaus, 1979), la détection de phonèmes (Dijkstra, Fieuws & Roelofs, 1995; Hallé, Chéreau & Seguí, 2000; Taft & Hambly, 1985) ou le *word blending* (Ventura, Kolinsky, Brito-Mendes & Morais, 2001). Or, puisque la capacité d'analyse phonologique explicite impliquée dans ces tâches est fortement liée à l'acquisition du code alphabétique (Morais, Cary, Alegria & Bertelson, 1979; Morais, Bertelson, Cary, & Alegria, 1986; Read, Zhang, Nie, & Ding, 1986), il n'est pas certain que les codes orthographiques à la base de ces jugements conscients jouent un rôle dans la reconnaissance des mots parlés (voir Morais, 1985; Morais & Kolinsky, 2001, pour des discussions). Cependant, quelques études ne sont pas atteintes par cette critique. Ainsi, Morais, Castro et Kolinsky (1991) ont étudié le phénomène de fusion phonologique (Cutting, 1975), qui se produit sous écoute dichotique et consiste à combiner erronément en un seul percept (e.g., "black") une consonne provenant d'un des stimuli (e.g., "lack") avec une autre venant de l'autre stimulus (e.g., "back"). En accord avec l'idée que l'orthographe influence l'identification des mots parlés, Morais *et al.* ont observé que le taux de fusions phonologiques chez les auditeurs lettrés est beaucoup plus faible lorsque l'orthographe du mot illusoire (e.g., "feliz") est incompatible avec le percept attendu (e.g., /fliʃ/ sur base de "liz"-"fiz"). Les études menées en français par Ziegler et

collaborateurs mènent à la même conclusion. Ziegler, Muneaux et Grainger (2003) ont trouvé que les latences de décision lexicale et de *shadowing* sont plus courtes pour les mots parlés à voisinage orthographique dense (e.g., "sèche") que pour ceux à voisinage orthographique épars (e.g., "sève"). Ziegler et Ferrand (1998) ainsi que Ziegler, Ferrand et Montant (2004) ont montré que la décision lexicale, la détection de rime, et le *shadowing* sont plus lents pour les mots parlés dont la rime peut s'écrire de plusieurs manières (e.g., "plomb") que pour ceux dont la rime ne s'écrit que d'une seule façon (e.g., "stage"; voir Ventura, Morais, Pattamadilok, & Kolinsky, 2004, pour des résultats similaires en portugais). Enfin, dans le paradigme d'amorçage auditif final, Chéreau, Gaskell et Dumay (sous presse) ont montré que le gain de traitement apporté par le recouvrement orthographique en décision lexicale (e.g., "gleam - dream" vs. "scheme - dream") ne diminue pas lorsque les *stimuli* sont comprimés, et que les latences sont conséquemment beaucoup plus courtes. L'orthographe semble donc être évoquée de manière automatique au cours de la reconnaissance des mots parlés.

2.4 Flux d'information dans le système de reconnaissance des mots parlés

La question du flux d'information dans le système de reconnaissance des mots parlés fait appel à la notion d'interactivité de processus et à son anti-thèse, à savoir la thèse d'autonomie computationnelle (cf. Fodor, 1983; Pylyshyn, 1984). Un processus *interactif* intègre dans le calcul de son signal de sortie des informations émanant d'étapes de traitement ultérieures dont il fournit l'entrée. A l'opposé, un processus *autonome* produit son signal de sortie sur base d'informations provenant uniquement d'étapes de traitement antérieures. Contrairement aux processus interactifs, les processus autonomes ne requièrent donc pas que le contenu des représentations qu'ils génèrent soit accessible à des processus subséquents avant que les dites représentations ne soient pleinement spécifiées —on parle, dans ce cas, d'architecture en cascade (cf. McClelland, 1979). Dans le cas du traitement lexical, le débat entre autonomie et interactivité porte essentiellement sur deux points: (1) l'activation des représentations lexicales influence-t-elle l'analyse perceptive —et donc pré-lexicale— des segments phonétiques contenus dans le signal? (2) l'identification du mot est-elle contrainte à son tour

par la structure syntaxique et sémantique (ou thématique) de la phrase dans laquelle le mot est enchâssé?

2.4.1 Effets du contexte lexical sur la perception?

En ce qui concerne le premier point, les démonstrations abondantes d'influences lexicales dans des tâches de jugement phonétique sont —naturellement— compatibles avec l'existence potentielle, comme dans TRACE, de connexions bidirectionnelles entre traitement lexical et traitement phonétique (cf. McClelland & Elman, 1986). En l'occurrence, un phonème est détecté plus rapidement dans un mot que dans un pseudomot (e.g., Cutler, Mehler, Norris, & Seguí, 1987; Rubin, Turvey, & van Gelder, 1976), et plus rapidement dans un pseudomot similaire à un mot que dans un pseudomot dissimilaire (Connine, Titone, Deelman, & Blasko, 1997; Wurm & Samuel, 1997). Par ailleurs, la catégorisation d'un phonème ambigu tend à préférer une interprétation lexicale du contexte; c'est ce qu'on appelle l'effet "Ganong" (e.g., /?/ est catégorisé comme /s/ dans /ki?/ mais comme /ʃ/ dans /fi?/; Connine & Clifton, 1987; Ganong, 1980; Pitt, 1995; Pitt & Samuel, 1995, sous presse). Ensuite, un phonème manquant est restauré "perceptivement" plus souvent dans un mot que dans un pseudomot (e.g., Samuel, 1981, 1996). Enfin, dans un mot, un phonème incompatible est détecté plus lentement lorsqu'il est similaire au phonème attendu que lorsqu'il n'y ressemble pas (Mirman, McClelland, & Holt, 2005). Cela dit, il faut admettre que tous ces effets peuvent aussi être expliqués en termes de biais lexical sur les mécanismes de décision (et non de perception) phonétique dans une architecture autonome, comme celle de Merge (Norris, McQueen, & Cutler, 2000).

Toutefois, et à côté d'arguments plus théoriques (cf. Bowers & Davis, 2004), deux ensembles de données empiriques suggèrent fortement que l'activation lexicale influence bel et bien le traitement perceptif des segments. Le premier ensemble de données est fourni par des études basées sur le phénomène de compensation pour la coarticulation, dans lequel l'identification d'une consonne plosive ambiguë à l'initiale de syllabe est influencée par l'identité de la fricative qui la précède (Mann & Repp, 1981; Repp & Mann, 1981, 1982). Ainsi, les auditeurs rapportent plus de plosives vélaires (/g/ ou /k/) après /s/ qu'après /ʃ/, et

inversement, plus de plosives dentales (/d/ ou /t/) après /ʃ/ qu'après /s/. Mann et Repp (1981) interprètent cet effet comme étant le complément perceptif naturel aux corrélats articulatoires de ces séquences, puisque lors de la production de ces dernières, le lieu d'articulation de la fricative (i.e., alvéolaire vs. palatal) affecte celui de la plosive subséquente. Afin de déterminer si ce phénomène pré-lexical de bas niveau est modulé en retour par l'activation lexicale, Elman et McClelland (1988) l'ont étudié dans une situation d'effet Ganong. Dans cette étude, une fricative perceptivement ambiguë (/?/), à mi-chemin entre /s/ et /ʃ/, est accolée à un contexte lexical permettant d'en influencer l'identification (e.g., "christma?" vs. "fooli?"), et immédiatement suivie d'un mot dont la plosive initiale varie le long d'un continuum (e.g., "tapes" vs. "capes"). En raison de l'effet Ganong, "christma?" conduit à entendre la fricative comme /s/, alors que dans "fooli?", elle est entendue comme /ʃ/. Elman et McClelland prédisaient que si une telle désambiguïsation résultait d'une authentique influence lexicale sur la perception des segments, l'identité de la fricative (/s/ ou /ʃ/), déterminée par le lexique, allait dès lors moduler la perception de la plosive /t/-/k/ subséquente via le phénomène de compensation pour la coarticulation. C'est exactement ce qu'ils ont trouvé: la catégorisation de la plosive montre plus de réponses /k/ lorsque le contexte lexical favorise la fricative /s/ (e.g., "christma?"), et plus de réponses /t/ lorsqu'il favorise la fricative /ʃ/ (e.g., "fooli?"). Par la suite, l'interprétation en termes d'effets haut-bas avancée par Elman et McClelland a été contestée, en raison du fait que, dans leurs items, le statut lexical était partiellement confondu avec la probabilité transitionnelle de la séquence /V+fricative/ (Shillcock, Lindsey, Levy, & Chater, 1992; Cairns, Shillcock, Chater, & Levy, 1995). Shillcock *et al.* (1992) ont montré qu'un modèle purement bas-haut ne disposant d'aucune représentation lexicale peut, sur seule base des probabilités transitionnelles stockées au niveau pré-lexical, simuler les résultats précédents (voir aussi Norris, 1993). En outre, Pitt et McQueen (1998) ont démontré, à l'aide de nonmots, que la compensation pour la coarticulation est fortement sensible aux probabilités transitionnelles. Cependant, puisque dans une des expériences publiées par Elman et McClelland (1988, Expérience 3), l'effet lexical émergeait alors même que les deux contextes se terminaient d'un point de vue acoustique par la même séquence /(C)V+fricative/, cette critique ne tient pas. L'obtention

récente par Samuel et Pitt (2003) et Magnuson, McMurray, Tanenhaus et Aslin (2003; voir McQueen, 2003, pour un commentaire) d'effets similaires dans des situations neutralisant ou renversant l'influence des probabilités transitionnelles confirme que l'évocation lexicale est à même de guider la compensation pour la coarticulation.

Le second ensemble de données empiriques est fourni par des expériences basées sur le phénomène d'adaptation sélective. Dans le cas de la parole, ce phénomène consiste en ce que la présentation répétée d'un prototype phonémique, appelé l'adaptateur, réduit la probabilité de classifier des exemplaires proches comme appartenant à la même catégorie (cf. Eimas & Corbit, 1973). Par exemple, après écoute répétée d'une séquence telle que /da/, la catégorisation d'un continuum /da/-/ta/ montre moins de réponses /da/, et vice versa lorsque l'adaptateur est /ta/. Afin de savoir si l'adaptation sélective est possible à partir d'un percept phonémique élaboré uniquement sur base de l'information lexicale, Samuel (1997) a examiné le phénomène en générant l'adaptateur au moyen du paradigme de restauration phonémique. En faveur de l'idée que le traitement du mot influence la perception phonétique, il a trouvé que le segment restauré du fait de l'activation lexicale (e.g., /b/ dans "exhi_ition" vs. /d/ dans "arma_illo") induit des effets d'adaptation. Plus récemment, et confortant encore davantage la position interactive, Samuel (2001) a obtenu des résultats similaires, au moyen d'un adaptateur illusoire /?/ créé par effet Ganong (e.g., /s/ dans "bronchiti?" vs. /ʃ/ dans "demoli?").

2.4.2 Effets du contexte phrastique sur l'identification des mots?

En ce qui concerne le second point, autrement dit l'influence potentielle des niveaux supérieurs d'analyse syntaxique et sémantique sur l'identification des mots, la vue classique s'inspire des études de Swinney (1979) et Tanenhaus et coll. (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Tanenhaus, Leiman, & Seidenberg, 1979) sur l'ambiguïté lexicale. Ces études montrent que lorsqu'une entrée lexicale est activée, tous les sens qui lui sont rattachés, qu'ils soient compatibles ou pas avec le contexte, deviennent automatiquement disponibles pour éventuellement être intégrés à la proposition en cours de construction. Selon cette conception, l'accès lexical est insensible aux influences d'ordre supérieur, et le contexte

phrastique ne joue de rôle que dans la sélection de la meilleure entrée parmi toutes celles qui sont évoquées. Toutefois, des travaux plus récents démontrent que le contexte est, dans certains cas, suffisamment puissant pour empêcher l'activation des représentations de mots incompatibles avec lui. Dans une situation où la structure syntaxique requiert que l'entrée soit appariée à un mot de classe fermée (ou mot fonctionnel; e.g., "would" dans "John said that he didn't want to do the job, but his brother would"), Shillcock et Bard (1993) n'ont trouvé aucune trace de l'activation d'un homophone de l'autre classe (ou mot à contenu; e.g., "wood"). Autre exemple: en utilisant la structure inversée "verbe-sujet" en néerlandais, Dahan et Tanenhaus (2004) ont montré que lorsque le verbe impose des contraintes sémantiques fortes sur le groupe nominal subséquent (cf. Altmann & Kamide, 1999; par exemple, le fait qu'un verbe comme "klimmen", *grimper*, requiert que le sujet soit nécessairement "animé"), seules les hypothèses lexicales compatibles en termes de leurs attributs sémantiques étaient susceptibles d'être évoquées. Dernier exemple: à l'aide de la détection de pause (voir Gaskell & Dumay, 2003; Mattys & Clark, 2002), Mattys, Pleydell-Pearce, Melhorn, & Whitecross (2005) ont comparé l'activité lexicale générée par des mots à point d'unicité (PU; voir chapitre 2, section 1.1) précoce ou tardif, selon que la phrase dans laquelle ces mots étaient enchâssés était neutre ou contraignante sur le plan sémantique (e.g., "The tribe had a gleeful deli[…]very" vs."The childbirth was a painful deli[…]very"). Conformément à l'idée que le contexte influence l'activation des compétiteurs lexicaux, ils ont trouvé que la différence d'activité lexicale était nettement plus réduite lorsque la phrase était contraignante. Contrairement à ce que suggèrent les thèses modularistes, des connexions haut-bas entre le niveau d'analyse syntaxico-sémantique et l'accès lexical pourraient donc être opérationnelles.

3 Interpréter l'énoncé et construire un modèle du discours

Processus inférenciel, la compréhension du message ne se réduit pas à simplement concaténer les sens attachés aux mots qui constituent l'énoncé. La phrase doit être analysée en termes de ses constituants, et le sens de chacun d'eux doit être évalué en tenant compte du contexte. Enfin, l'ensemble de la phrase doit être interprété sémantiquement et la valeur communicative de celle-ci par rapport au discours doit être déterminée.

L'interprétation sémantique de l'énoncé se fait de manière incrémentale et presque immédiate. L'auditeur n'attend pas la fin de la proposition ou de la phrase pour générer une représentation propositionnelle; au contraire, le processus de compréhension paraît traquer les mots au fur et à mesure qu'ils arrivent, quitte à tomber momentanément dans une interprétation inattendue, voire peu plausible. L'existence du phénomène de *garden-path* est sans doute l'argument le plus solide en faveur de cette idée. Dans des phrases telles que "The horse raced past the barn fell", "raced", lorsqu'il est un verbe d'action (Spivey-Knowlton & Sedivy, 1995), est systématiquement interprété de façon erronée comme étant le verbe principal, avec comme conséquence que lorsqu'intervient le verbe suivant (e.g., "fell"), la phrase devient totalement incompréhensible (cf. Bever, 1970). Autre argument en faveur de l'idée que l'interprétation sémantique est un processus incrémentale, le verbe impose des contraintes sémantiques fortes sur la reconnaissance de l'argument subséquent alors que celui-ci n'a pas encore été entendu (Altmann & Kamide, 1999). Cette capacité à élaborer d'emblée une interprétation de l'énoncé n'est évidemment possible que parce que le traitement lexical est rapide, robuste et efficace. Les caractéristiques mêmes des processus intégratifs suggèrent donc qu'il est peu probable que la segmentation découle de la reconnaissance d'unités prosodiques plus larges que le mot.

L'un des débats dans le domaine concerne le nombre d'analyses menées simultanément, et touche donc à la flexibilité du système. Selon les modèles sériels (Frazier, 1979, 1987, 1989; Frazier & Clifton, 1996), l'analyseur syntaxique ne peut jamais construire qu'une seule interprétation de l'énoncé. En revanche, les modèles interactifs (e.g., McDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994) postulent que plusieurs interprétations sont élaborées simultanément et sont en compétition les unes avec les autres.

Un autre débat porte sur la nature de l'information syntaxique utilisée dans les premiers stades de l'analyse de la phrase. Pour certains chercheurs (e.g., Frazier, 1979, 1987, 1989), la structure grammaticale doit d'abord être construite, afin de pouvoir étayer toute interprétation sémantique, et seules des connaissances grammaticales générales seraient employées. Pour d'autres (e.g., Altmann, 1999; McDonald *et al.*, 1994; Tanenhaus, Garnsey, & Boland, 1990; Tanenhaus, Boland, Mauner, & Carlson, 1993), influencés à la fois par la notion de

satisfaction de contraintes et par les théories lexicalistes (e.g., Pustejovsky, 1995), l'information utilisée, beaucoup plus riche, inclurait à la fois des éléments grammaticaux spécifiques, tels que la connaissance des structures d'arguments liés à un verbe donné, et des éléments extra-grammaticaux, comme, par exemple, la fréquence d'usage relative de diverses constructions.

Pour clore cette section, notons que dans le cas de la modalité auditive, on peut s'attendre à ce que l'incertitude du système quant à la segmentation lexicale de l'énoncé interagisse de façon remarquable avec l'ambiguïté de sa structure syntaxique.

Chapitre 2 Les mécanismes de la segmentation lexicale: théories et éléments empiriques

Deux types de solutions ont été proposées pour rendre compte de la segmentation en mots. Les premières sont pré-lexicales et exploitent certaines propriétés du signal liées aux frontières de mots. Ces propriétés sont souvent désignées par le terme "indices de segmentation". Les secondes sont lexicales et envisagent la segmentation comme une conséquence indirecte de la reconnaissance des mots. En ce sens, les solutions du deuxième type sont générales, indépendantes de la langue et applicables par défaut, puisque, en principe, elles considèrent qu'il est possible d'extraire un mot du continuum de parole même en l'absence de marquage acoustique fiable des frontières lexicales. Comme nous le verrons, ces deux approches ne sont pas incompatibles. Dans ce qui suit, nous les envisageons tour à tour et revoyons de manière détaillée les éléments empiriques qui attestent de leur pertinence. Nous commencerons par l'approche lexicale.

1 Approche lexicale: la segmentation comme produit de la reconnaissance des mots

Les solutions au problème de la segmentation qui se fondent sur nos connaissances lexicales se répartissent en gros en deux classes: celles qui partent de l'idée que les mots parlés sont reconnus de manière strictement séquentielle (Cole & Jakimik, 1978, 1980a; Marslen-Wilson & Welsh, 1978); et celles qui exploitent le processus de compétition active entre candidats lexicaux évoqués par une même portion du signal (McClelland & Elman, 1986; Norris, 1994). Dans les deux cas, le découpage des énoncés en leurs constituants lexicaux ne relève pas d'une procédure spécifique; au contraire, il découle implicitement de la reconnaissance des mots (Reddy, 1976).

1.1 La reconnaissance séquentielle des mots

L'idée que le traitement lexical suivrait rigoureusement le déploiement de l'information acoustique dans le temps, et serait, donc, strictement séquentiel, est l'apanage des premiers

modèles de reconnaissance propres à la modalité auditive (Chodorow, 1979; Cole, 1973; Cole & Jakimik, 1978, 1980a; Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980; Marslen-Wilson, Tyler, & Seidenberg, 1978). Selon ces modèles, dont le plus connu est la théorie initiale de la Cohorte (de Marslen-Wilson et coll.), chaque début de mot activerait en parallèle tous les mots du lexique commençant par cette séquence (ces mots constituant la "cohorte initiale" du stimulus); ensuite, les candidats incompatibles avec l'information sensorielle subséquente ou les contraintes contextuelles du message seraient désactivés jusqu'à ce qu'il ne reste qu'un mot —idéalement, celui qui est présenté.

Une propriété importante de ces modèles est que ceux-ci n'utilisent que la quantité d'information nécessaire à l'identification du mot. Dès lors, le point de reconnaissance d'un mot correspond de manière linéaire à son *point d'unicité* (PU; point à partir duquel il est le seul, dans le lexique, à pouvoir rendre compte de la totalité de l'information sensorielle). Sur base d'un mécanisme de reconnaissance aussi optimal et aussi précoce, l'auditeur pourrait donc anticiper la fin du mot actuel, et, partant, prédire à quel endroit débute le mot suivant.

L'existence d'une composante séquentielle dans le traitement des mots parlés a reçu de nombreuses confirmations empiriques, notamment sous la forme des effets directionnels dans l'activation des représentations lexicales; ainsi, ces représentations sont évoquées plus facilement par leur début que par leur fin (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Cole & Jakimik, 1980b; Gaskell & Dumay, 2003; Magnuson, Tanenhaus, Aslin, & Dahan, 2003; Marslen-Wilson & Zwitserlood, 1989; Nooteboom, 1981; Zwitserlood, 1989). L'existence d'une telle composante séquentielle est également appuyée par les effets du PU: les mots à PU précoces sont généralement reconnus plus rapidement que les mots à PU tardifs (e.g., Frauenfelder, Seguí, & Dijkstra, 1990; Marslen-Wilson, 1987; Mattys & Clark, 2002; Radeau, Morais, Mousty, & Bertelson, 2000; Radeau, Mousty, & Bertelson, 1989; Radeau & Morais, 1990; Taft & Hambly, 1986; voir Mattys, 1997, pour un état de la question).[3]

[3] Radeau *et al.* (2000) ont récemment mis en doute la pertinence des mécanismes sous-jacents aux effets du PU dans le traitement de la parole continue. Dans leur étude, l'effet de la position du PU en décision de genre et *shadowing* était examiné en fonction du taux d'articulation des stimuli, lequel pouvait être "lent" (2.2 syllabes (syll)/s), "moyen" (3.6 syll/s) ou "rapide" (5.6 syll/s). Aucun effet de la position du PU n'était observé à débit rapide. En revanche, à débit moyen, un effet émergeait en décision de genre, tandis qu'à débit lent, un effet était présent dans les deux tâches. Les auteurs ont conclu qu'étant donné la forte similarité entre le taux d'articulation conversationnel et celui de la condition de débit rapide, les effets du PU devaient être imputables à des

Cependant, des expériences de *gating* ont montré que les auditeurs ne sont pas toujours capables d'identifier les mots sans que l'information subséquente ne leur soit donnée (Bard, Shillcock, & Altmann, 1988; Grosjean, 1985; Pollack & Pickett, 1963; Pickett & Pollack, 1963). Par exemple, à l'aide d'extraits conversationnels présentés par pas d'un mot, Bard *et al.* (1988) ont trouvé que 21% des identifications correctes impliquent le contexte subséquent. Les plus fréquentes de ces identifications tardives touchent les mots courts, les mots fonctionnels et les mots situés en début d'énoncé.

L'exploration de la structure des bases de données lexicales donne du sens aux résultats de ces expériences de *gating*, en mettant en évidence que beaucoup de mots ne divergent des autres mots du lexique qu'après leur dernier phonème. Par exemple, sur base d'un dictionnaire d'anglais américain de 20.000 mots, Luce (1986b) a montré que si l'on tient compte de la fréquence d'occurrence des mots, 38% d'entre eux ne sont pas uniques à leur segment final. Autrement dit, selon cette analyse, plus d'un tiers des mots de l'anglais constituent la portion initiale d'un mot plus long. Des probabilités similaires ont été obtenues pour le néerlandais par Frauenfelder et Peeters (1990), et pour le français par Radeau *et al.* (2000). Dans le cas du français, Radeau *et al.* indiquent qu'en moyenne, le PU n'est situé dans le mot que si le mot contient au moins cinq phonèmes. Par ailleurs, McQueen, Cutler, Briscoe et Norris (1995) ont trouvé que même si le décompte des mots enchâssés est restreint à ceux qui ont la même structure syllabique que le mot enchâssant, 57.5% des mots polysyllabiques de l'anglais ont au moins un mot enchâssé pour première syllabe. Conformément aux résultats de Frauenfelder et Peeters (1990) pour le néerlandais, McQueen *et al.* observent que si la fréquence

---

mécanismes de traitement optionnels n'ayant que peu de rapports avec ceux qui sont impliqués dans la reconnaissance de la parole conversationnelle. Cette conclusion est cependant trop forte. La similitude entre le taux d'articulation des mots isolés dans la condition rapide et celui de la parole enchaînée ne permet pas d'inférer que les mécanismes de traitement à l'œuvre dans cette condition soient ceux utilisés en situation normale. Le taux d'articulation étant lié par une fonction asymptote à la longueur phonologique de l'énoncé (Bartkova, 1991; Fónagy & Magdics, 1960; Gaitenby, 1965; Malécot, Johnston, & Kizziar, 1972; Pollack & Pickett, 1964), un mot est plus court lorsqu'il se trouve dans une phrase que lorsqu'il est prononcé isolément, signifiant qu'un taux d'articulation normal pour de la parole enchaînée pourrait être beaucoup trop rapide pour des mots produits isolément. Sans une estimation de la durée que prennent les substantifs trisyllabiques utilisés par Radeau *et al.* lorsqu'ils sont prononcés de manière isolée au cours d'une conversation, il est dès lors prématuré de conclure qu'il n'y a pas d'effet du PU à un taux d'articulation normal. L'adjonction d'un contexte de phrase aux stimuli utilisés pourrait d'ailleurs permettre l'émergence d'un effet du PU à débit rapide.

d'enchâssement est élevée pour toutes les positions dans le mot, elle est systématiquement la plus élevée à l'initiale.

Au regard de ces statistiques lexicales, la stratégie de segmentation par anticipation, nécessaire aux modèles strictement séquentiels, devrait donc être d'une efficacité très faible, puisque, dans un nombre substantiel de cas, le système de traitement ne pourrait rejeter la possibilité d'un compétiteur plus long avant le début du mot suivant (voir Norris, 1990, et Gaskell & Marslen-Wilson, 1997, pour des démonstrations connexionnistes de ce phénomène).

Les résultats obtenus récemment au moyen du *garden-path* lexical par Davis, Marslen-Wilson et Gaskell (2002; voir ce chapitre, section 2.1.2.3) obligent toutefois à nuancer cette conclusion. Lorsque le système de traitement est confronté à un cas d'enchâssement initial potentiel (e.g., "cap tucked"), il entretient au-delà de la fin du mot la possibilité d'un compétiteur plus long (e.g., "captain"; cf. Content & Sternon, 1994; Tabossi, Burani, & Scott, 1995). Toutefois, comme l'ont démontré Davis *et al.*, des indices acoustiques présents dans le signal permettent au système de distinguer bien avant la fin de la première syllabe si c'est le mot court ou son compétiteur plus long qui est articulé. L'ambiguïté résultant de l'occurrence naturelle dans le lexique de mots enchâssés à l'initiale ne paraît donc pas aussi problématique que ne laissent supposer les statistiques lexicales.

Par conséquent, s'il est clair que la stratégie de segmentation séquentielle ne peut être utilisée de manière systématique, elle n’est certainement pas seulement exploitée pour les seuls mots du lexique, longs et rares, qui divergent sur le plan phonémique des autres mots avant leur fin.

1.2 Le processus de compétition lexicale

Les modèles connexionnistes basés sur l'hypothèse d'alignement exhaustif, tels que TRACE (McClelland & Elman, 1986) et Shortlist (Norris, 1994), offrent une solution générale qui permet de concevoir le système de traitement comme recourant au contexte droit dans son évaluation du découpage de l'énoncé. L'alignement exhaustif renvoie au fait que, dans ces théories, les recherches lexicales ne sont pas restreintes à certains point de

segmentation; au contraire, chaque nouveau segment phonémique est considéré comme un début de mot potentiel. Dans TRACE et Shortlist, de multiples candidats lexicaux, pouvant débuter à des moments différents du temps, sont en compétition active pour l'appropriation des portions du signal avec lesquelles ils sont compatibles. Dans ces modèles, la compétition lexicale est instanciée par l'existence de connexions inhibitrices bi-directionnelles entre les représentations; la force de l'inhibition entre deux unités lexicales est proportionnelle au nombre de segments phonémiques par lesquels elles se chevauchent dans le temps.

Dans TRACE, réseau d'activation interactive (cf. McClelland & Rumelhart, 1981), la compétition est implémentée dans sa dimension temporelle par une duplication complète du réseau, avec ses trois couches d'unités —traits phonétiques, phonèmes et mots—, à chaque début de mot potentiel.[4] Tous les mots du lexique sont à tout moment en compétition constante pour la reconnaissance. Grâce à la dynamique de l'activation interactive, et, en particulier, grâce aux connexions inhibitrices entre les représentations lexicales, le système converge rapidement vers une représentation optimale de l'énoncé. Celle-ci rend compte de la totalité du signal, et les mots qui la constituent ne se chevauchent par aucun segment. Pour reprendre un exemple donné par McClelland et Elman (1986, p. 63, Figure 28), lorsque TRACE reçoit une séquence telle que /bɑːti/ (bar tea), l'unité correspondant au mot "art" entre en compétition avec celles correspondant aux mots "bar" et "tea". Mais comme les mots "bar" et "tea" ne se recouvrent pas, et qu'à eux seuls, ils rendent compte de l'entièreté de l'énoncé, ils inhibent conjointement le mot "art", et constituent au final la réponse du système (pour une exploration systématique de la segmentation dans TRACE, voir Frauenfelder & Peeters, 1990). Toutefois, d'un point de vue computationnel, TRACE est peu économique. Comme il ouvre en principe la compétition à n'importe quel mot du lexique à tout moment du temps, le nombre de connections inhibitrices qui doivent être dupliquées à chaque début de mot potentiel est énorme (pour une démonstration simple, voir Norris, 1994, pp. 198-199). Dans les modèles qui ont suivi, la question de l'économie computationnelle a donc été prise en compte.

[4] Dans les faits, TRACE recrée une copie complète de son réseau trois fois par phonème, de manière à pouvoir simuler les effets de coarticulation.

Shortlist (Norris, 1994) résout le problème lié à la duplication de l'entièreté du réseau en séparant le processus de génération des candidats à la reconnaissance du processus de compétition lexicale. Le modèle fonctionne en deux temps. Dans un premier temps, un ensemble de candidats potentiels est généré de manière strictement "bas-haut", sur base de leur degré d'appariement avec le signal. Contrairement à ce qui se passe pour TRACE, dans Shortlist, le nombre maximal de candidats potentiels est fixé au préalable. En outre, la représentation d'entrée active les candidats compatibles avec elle, mais elle inhibe, d'un autre côté, tous ceux qui sont incompatibles (Frauenfelder, Scholten, & Content, 2001; Marslen-Wilson, 1993). Dans un second temps, les candidats lexicaux en lice sont câblés les uns aux autres de façon à constituer un réseau d'activation interactive, qui fonctionne exactement de la même manière que la couche lexicale de TRACE. C'est à ce niveau que les candidats lexicaux retenus entrent en compétition pour la reconnaissance.

La Figure 2 illustre la manière dont émerge la segmentation de la dynamique de l'activation interactive; elle présente la réponse de Shortlist face à la séquence /ʃɪpɪŋkwaɪəri/ ("ship inquiry"). Dans cette séquence, l'interprétation de la première partie de l'énoncé est totalement dépendante des phonèmes qui suivent.

Comme on peut le voir, le compétiteur chevauchant "shipping" est d'abord activé au détriment du premier mot "ship". Néanmoins, lorsque s’accumule l'information révélant l'identité du second mot "inquiry", et que se renforce l'action inhibitrice conjointe de "ship" et de "inquiry" sur le compétiteur chevauchant "shipping", "ship" perce à son tour et le réseau se stabilise sur la segmentation correcte. Les modèles d'activation interactive peuvent donc segmenter sans difficultés des énoncés même ambigus sur une portion importante du signal.

Vu la performance de ces modèles, la compétition entre de multiples candidats lexicaux offre une solution très robuste au problème de la segmentation. Dans la section qui suit, nous offrons un aperçu des résultats expérimentaux qui suggèrent qu’un tel mécanisme existe chez l'humain.

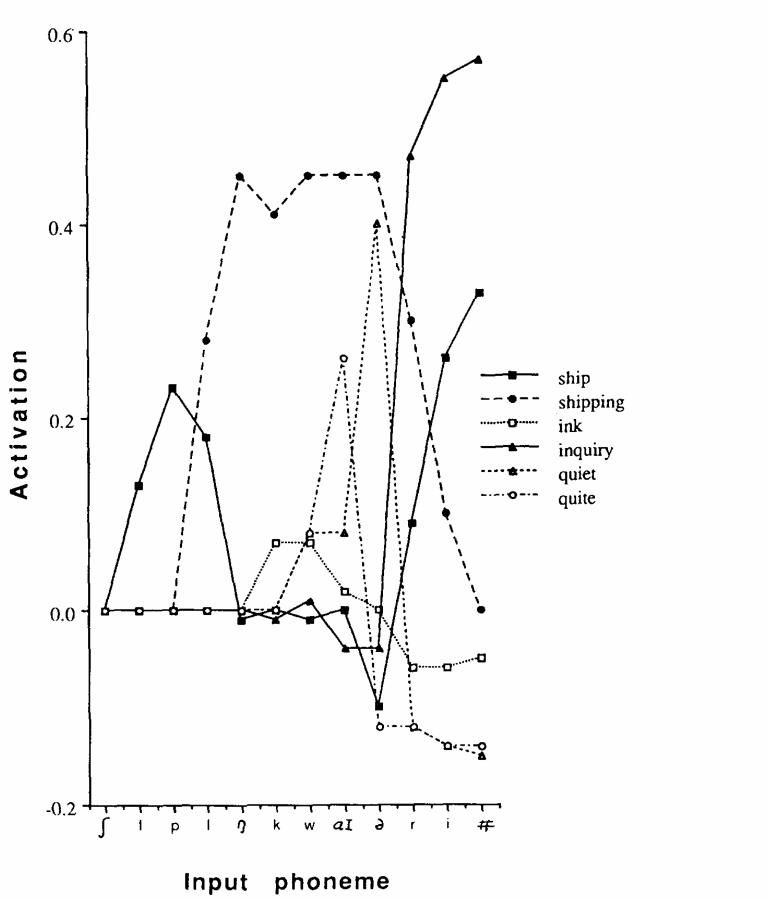

Figure 2. Niveaux d'activation des mots inclus dans la liste des candidats lexicaux produite par Shortlist au cours de la présentation de la séquence "ship inquiry". Dans cette simulation, le nombre maximal de candidats est fixé à 30 (reproduit à partir de Norris, 1994, Figure 8, avec la permission de l'auteur).

1.2.1 Effets de compétition lexicale

Un nombre important d'études suggèrent que les candidats lexicaux sont activés de manière multiple et automatique en tout point du signal (Marslen-Wilson 1987, 1993; Zwitserlood, 1989, pour les compétiteurs cohortiens; Isel & Bacri, 1999; Luce & Cluff, 1998; Shillcock, 1990; Vroomen & de Gelder, 1997a, pour les mots enchâssés en finale d'autres mots; Davis *et al.*, 2002; Gow & Gordon, 1995; Tabossi *et al.*, 1995, pour les compétiteurs chevauchant les frontières lexicales; voir aussi Gaskell & Dumay, 2003; Dumay & Gaskell, 2005, sous presse; Dumay, Gaskell, & Feng, 2004; Magnuson *et al.*, 2003, pour des démonstrations utilisant des paradigmes d'apprentissage). Toutefois, le fait que les candidats

soient activés ne suffit pas à établir qu'ils sont effectivement en compétition active les uns avec les autres.

L'inhibition de la reconnaissance d'un mot du fait de l'activation d'un autre mot constitue la preuve la plus directe du phénomène de compétition active. Des effets d'inhibition dans la reconnaissance des mots ont été mis en évidence essentiellement dans deux types d'expériences: celles qui modifient l'état d'activation des compétiteurs par le biais de l'amorçage formel, et celles qui jouent sur la présence ou le nombre de compétiteurs enchâssant ou chevauchant.

Les expériences d'amorçage formel ont envisagé les liens de similarité entre les mots soit en termes de ressemblance acoustico-phonétique, soit en termes de recouvrement phonémique. Goldinger et coll. ont montré que la reconnaissance d'un mot (e.g., "bull") peut être inhibée par la présentation préalable d'un mot qui lui est phonétiquement similaire (e.g., "veer"), dans des tâches telles que l'identification perceptive et la décision lexicale dans le bruit (Goldinger, Luce, & Pisoni, 1989; Goldinger, Luce, Pisoni, & Marcario, 1992; voir aussi, Luce *et al.*, 1990), et plus récemment le *shadowing* (Luce *et al.*, 2000). Etant donné la subtilité du lien entre l'amorce et la cible, l'interprétation préférée des auteurs est que la présentation préalable de l'amorce permet à celle-ci d'être activée plus efficacement par la cible, et, partant, de retarder le traitement de cette dernière.

Des effets d'inhibition ont également été mis en évidence au moyen de l'amorçage phonologique initial (e.g., "cube-cure"). Bien que les premières études examinant cet effet aient produit des résultats assez peu cohérents (voir Radeau, Morais, & Seguí, 1995, pour un état de la question), des études plus récentes contrôlant à la fois le taux de recouvrement et la contribution de facteurs stratégiques montrent que les latences d'identification sont plus longues lorsque la cible partage avec l'amorce un nombre substantiel de phonèmes[5] (Dufour & Peereman, 2003a; Hamburger & Slowiaczek, 1996; Radeau *et al.*, 1995; Slowiaczek & Hamburger, 1992; voir Dumay *et al.*, 2001, pour un état de la question). En outre, dans une étude récente, Dufour et Peereman (2003b) ont trouvé que l'influence inhibitrice de l'amorce

[5] Les résultats de Dufour et Peereman (2003a) montrent d'ailleurs que pour que l'amorce inhibe la cible, il faut qu'elle ait en commun avec elle tous ses phonèmes sauf le dernier.

n'est visible que lorsque la cible (et partant l'amorce) ne possède qu'un nombre restreint de compétiteurs; lorsque les compétiteurs sont nombreux, l'amorce n'est que faiblement réactivée, sans doute en raison de l'inhibition qu'elle subit elle-même, ce qui explique qu'elle ne peut donc à son tour inhiber la cible.

Contrairement aux précédentes, les expériences examinant les effets des compétiteurs enchâssant ou chevauchant concernent plus directement la segmentation. Au moyen de la tâche de *word spotting*, McQueen, Norris et Cutler (1994) ont cherché à établir si la détection d'un mot enchâssé dans une séquence sans signification était affectée négativement par le fait que la séquence enchâssante soit le début d'un mot. Pour les mots enchâssés à l'initiale, les auteurs n'ont pas trouvé de trace fiable de l'activation du compétiteur plus long, tout au plus une légère tendance était-elle observée. Par exemple, "sack" n'était pas statistiquement plus difficile à détecter dans "sacrif", compatible avec "sacrifice", que dans "sackrek", qui n'est pas le début d'un mot. En revanche, pour les mots enchâssés à la finale, la compatibilité lexicale de la séquence porteuse inhibait significativement la reconnaissance du mot. Ainsi, "mess" était détecté plus lentement et moins souvent dans "domess", compatible avec "domestic", que dans "nemess", qui n'est compatible avec aucun mot. De manière similaire en néerlandais, Cutler et van Donselaar (2001) ont trouvé qu'il est plus difficile de détecter "zee" dans "muze" compatible avec "muzeum" que dans "luze".

Au moyen de l'amorçage sémantique intermodal, Vroomen et de Gelder (1997a) ont trouvé qu'un mot enchâssé à l'initiale (e.g., "vel") facilitait le traitement d'une cible visuelle lorsque la séquence porteuse était un nonmot (e.g., "velk"), mais pas lorsqu'elle était un mot (e.g., "velg"); cela corrobore la tendance observée par McQueen *et al.* à l'aide de séquences compatibles et de mot enchâssés à l'initiale.

En outre, deux études, réalisées par Vroomen et de Gelder (1995) et Norris, McQueen et Cutler (1995), suggèrent que plus un mot active de compétiteurs chevauchants, plus sa reconnaissance est inhibée. Ainsi, Vroomen et de Gelder ont trouvé qu'en répétition intermodale, l'amorçage facilitateur produit par un mot enchâssé à l'initiale est d'autant plus faible que le nombre de compétiteurs débutant avec la consonne finale de la cible est grand (e.g., "melk" dans "melkeum" vs. "melkaam"). A l'aide d'une tâche de *word spotting*, Norris *et*

al. (1995) ont trouvé qu'il est d'autant plus difficile de détecter un mot à l'initiale qu'il y a de compétiteurs (activés via MSS) débutant sur le dernier phonème de la cible (e.g., "stamp" in "stampidge" vs. "stampedge" comparé à "mint" in "mintaup" vs "mintep"). Par ailleurs, dans les deux études, l'absence de chevauchement annihile l'influence du nombre de compétiteurs (e.g., "bel" dans "belkeum" vs. "belkaam").

Dans l'ensemble, ces résultats montrent que les candidats alignés sur le début de l'énoncé aussi bien que les candidats non-alignés entrent en compétition les uns avec les autres dans le but de s'approprier les mêmes portions du signal. La compétition active entre les hypothèses lexicales pourrait donc jouer un rôle déterminant dans la segmentation.

1.2.2 Limites de la compétition lexicale en tant que solution à la segmentation

Comme on l'a vu, la solution à la segmentation qu'offre la compétition lexicale est applicable par défaut, et c'est ce qui fait son intérêt. En l'absence totale de marqueurs des frontières, elle permet au système de résoudre les cas d'enchâssement potentiel en postposant les décisions jusqu'à l'occurrence d'une portion de parole incompatible avec l'une ou l'autre des hypothèses lexicales encore entretenues. Une des implications de l'engagement tardif est que l'information temporairement non-segmentée doit être stockée dans un *buffer* acoustique ou en mémoire de travail en attendant la levée de l'ambiguïté (voir Mattys, 1997, pour une discussion). Par conséquent, si cette solution permet d'éviter bon nombre d'erreurs grossières de segmentation, elle sollicite davantage les capacités de stockage à court-terme du cerveau humain, lesquelles ne sont pas illimitées.

Par ailleurs, la langue comporte de nombreux cas d'énoncés homophoniques (du point de vue phonémique; voir El-Bèze, 1996, pour des statistiques sur le français), difficile à segmenter uniquement sur base des connaissances lexicales sans commettre d'erreur. Etant donné que notre système de perception ne semble avoir aucune difficulté à traiter ces énoncés, il est nécessaire de postuler l'exploitation d'indices dérivés du signal.

2 Approche prélexicale: l'exploitation d'indices de segmentation

Les énoncés se caractérisent en général par de multiples phénomènes d'ordre acoustique, phonétique, ou encore phonologique, induits par la présence d'une frontière de mots ou tout simplement liés statistiquement à cette frontière (voir Trubetzkoy, 1939, pour une première taxinomie). Certains de ces phénomènes mettent en jeu les segments touchant directement à la frontière, ou segments dits "limites", tandis que d'autres interviennent dans un voisinage beaucoup plus large. L'exploitation par l'auditeur de ces corrélats des frontières lexicales pourrait dès lors lui rendre la tâche de segmentation nettement plus aisée. Plus spécifiquement, elle pourrait lui permettre de déterminer à quels endroits du signal déclencher préférentiellement les recherches lexicales, et donc contraindre le processus de reconnaissance des mots.

Les trois sections suivantes présentent les données de la littérature et les principales hypothèses concernant l'utilisation des indices de segmentation. L'état de la question s'organise en fonction de la nature de l'information utilisée: indices acoustico-phonétiques, indices syllabiques, indices phonotactiques et probabilités transitionnelles, et indices rythmiques. Etant donné l'intérêt particulier que nous porterons, dans la partie expérimentale, au rôle joué par les indices acoustico-phonétiques, la section qui les concerne sera la plus importante.

2.1 Indices acoustico-phonétiques

L'idée que les frontières de mots pourraient être localisées sur base de variations qualitatives et/ou quantitatives dans la réalisation des segments a une longue histoire (e.g., Jones, 1931; Sweet, 1877). Toutefois, c'est seulement depuis les travaux de Lehiste (1960, 1961, 1964) qu'on a commencé à examiner cette hypothèse en combinant des mesures sur la perception des énoncés à l'expertise phonétique.

Les recherches dans le domaine sont basées presque toujours sur des énoncés à structure lexicale ambiguë; deux cas de figure ont été particulièrement examinés: (1) le cas où un phonème, que l'on qualifiera de segment "pivotal", peut être assigné soit au mot qui précède, soit au mot qui suit, comme dans une paire minimale telle que "danse errante" vs. "dent

serrante"; et (2) le cas où deux syllabes appartiennent soit au même mot, soit à des mots différents, comme dans une paire minimale telle que "car fourgué" vs. "carrefour gai". Dans le premier cas, l'interprétation de l'énoncé dépend crucialement de la structure syllabique; en revanche, dans le second, cette dimension n'est pas nécessairement impliquée.

Par souci de clarté, nous aborderons les deux cas séparément. Pour chacun d'eux, nous donnerons un aperçu des phénomènes phonétiques susceptibles d'être utilisés pour distinguer les alternatives, puis nous reverrons les données expérimentales mesurant la sensibilité des auditeurs à ces indices de segmentation. En particulier, nous distinguerons les études examinant les choix de segmentation, forcément explicites, des auditeurs, et les études examinant la segmentation au cours de la reconnaissance des mots. Comme nous le verrons, les données sur l'anglais dominent la littérature, mais certaines des observations les plus pertinentes proviennent d'études sur le néerlandais et sur le français.

2.1.1 Assignation d'un phonème au mot qui précède ou à celui qui suit

2.1.1.1 Aperçu des variations phonétiques

D'une manière générale, les alternatives d'une paire minimale telle que "dan<u>s</u>e errante" vs. "dent <u>s</u>errante" se distinguent essentiellement par les propriétés du segment pivotal. D'une part, ce segment tend à être d'une durée acoustique plus longue à l'initiale du second mot qu'en finale du premier. L'allongement du phonème initial est un phénomène très robuste, attesté dans de nombreuses langues. Celles-ci incluent notamment le français (Content, Bagou, Frauenfelder, & Fougeron, 2004; Dejean de la Bâtie, 1993; Fougeron, Bagou, Stefanuto, & Frauenfelder, 2002; Spinelli, McQueen, & Cutler, 2003; Vaissière, 1977, 1989), l'allemand (Kohler & Hartcastle, 1974; Kohler, 1976), l'anglais (Hoard, 1966; Klatt, 1974; Lehiste, 1960; Oller, 1973; Umeda, 1977), le néerlandais (Nooteboom, 1972; Quené, 1992, 1993), l'italien (Nespor, 1977), le tchèque (Lehiste, 1964), et le suédois (Carlson, Granström, Lindblom, & Rapp, 1972; Lindblom & Rapp, 1973).

D'autre part, l'allophone initial se caractérise par des propriétés articulatoires qui, combinées à l'allongement de la durée de ce segment, tendent à renforcer celui-ci et à le

distinguer des allophones non-initiaux (voir Fougeron, 1998, pour un état de la question). La marge de variation possible dépend évidemment de l'identité du phonème en question. Les variations les plus remarquables incluent l'aspiration systématique des plosives sourdes à l'initiale de syllabe, ce qui a pour conséquence de dévoiser partiellement ou totalement la voyelle, semi-voyelle ou liquide subséquente, notamment en anglais (e.g., Christie, 1974; Kahn, 1976; Lehiste, 1960) et en suédois (Garding, 1967); la distinction, en anglais, entre les deux allophones de /l/, l'un, qualifié de "*light*", attesté uniquement en début de syllabe, et l'autre, qualifié de "*dark*", attesté seulement à la coda (e.g., Lehiste, 1960; Trubetzkoy, 1939; Umeda & Coker, 1975); ou encore, la glottalisation —friction laryngée voire, plus rarement, coup de glotte— devant les voyelles à l'initiale de mot en anglais (e.g., Gimson, 1980; Lehiste, 1960; Umeda 1978), en allemand (Kohler, 1994), en tchèque, ou encore en finnois (Lehiste, 1964).[6]

D'autres variations, plus subtiles, semblent elles aussi avoir une fonction de renforcement. Ainsi, la position initiale dans la syllabe ou dans le mot est associée à un accroissement de l'activité musculaire labiale (e.g., Bonnot, Chevrie-Muller, Arabia-Guidet, Maton, & Grenier, 1986; Straka, 1963, pour le français; Fromkin, 1965, pour l'anglais). Elle se caractérise par une position du voile du palais plus élevée (e.g., Benguerel, 1977, pour le français; Krakow, 1989, 1993, pour l'anglais; Fujimura, 1990, pour le japonais). Elle est marquée par un contact et/ou par une pression linguo-palatale plus importante —car la langue est plus haute (e.g., Fougeron, 2001; Rousselot, 1901; Straka, 1963, pour le français; Byrd, 1994, 1996; Browman & Goldstein, 1995; Keating & Wright, 1994; Keating, Wright, & Zhang, 1999, pour l'anglais). Enfin, cette position se différencie par une coordination spécifique des articulateurs (e.g., Krakow, 1989, 1999), et par un recouvrement temporel moins important et plus stable entre les segments (e.g., Byrd, 1996).

Outre les variations temporelles et articulatoires sur le segment pivotal, des différences de durée ont aussi été observées sur la voyelle qui précède la frontière, notamment pour le français et pour l'anglais. En français, cette voyelle tend à être plus longue et d'un premier

[6] Etant donné la nature agglutinante de l'allemand, dans cette langue, la glottalisation des voyelles n'est pas un marqueur des frontières de mots mais un marqueur des frontières de morphèmes.

formant plus élevé en position non-limite (VC#[7]) qu'en position limite (V#), mais la nature systématique de ces deux phénomènes reste à déterminer (Content *et al.*, 2004; Fougeron *et al.*, 2002). Par contre, les résultats obtenus en anglais sont peu cohérents (Barry, 1981; Crystal & House, 1990; Hoard, 1966; Lehiste, 1960).

Mais avant d'aller plus loin, une remarque s'impose. Les études qui mettent en évidence l'allongement et/ou le renforcement articulatoire du segment à l'initiale de mot sont presque toutes basées exclusivement sur des monosyllabes, où, par définition, les frontières de mots se confondent avec les frontières de syllabes. Par conséquent, ces études ne permettent pas de dire si les phénomènes observés sont liés spécifiquement à la présence de la frontière lexicale. Par ailleurs, les quelques études comparant les débuts de mots aux débuts de syllabes (e.g., "a tune" vs. "attune") en termes d'allongement ou de renforcement articulatoire (ici au moyen de l'électropalatographie) n'ont pas trouvé de différence fiable et systématique (Krakow, 1989; Fougeron *et al.*, 2002; Keating *et al.*, 1999). La spécificité des débuts de mots dans la chaîne parlée pourrait donc avant tout refléter leur statut d'attaques syllabiques.

2.1.1.2 Etudes perceptives basées sur les choix de segmentation explicites

Dans une étude portant sur la description des allophones positionnels, Lehiste (1960) a enregistré plusieurs locuteurs anglophones lisant des paires d'énoncés, dont la plupart (19 sur 25) ne différaient l'un de l'autre que par l'assignation lexicale d'un segment (e.g., play taught" vs. "plate ought"). La sensibilité des auditeurs à d'éventuelles différences était examinée à l'aide d'un paradigme dans lequel les participants devaient déterminer l'ordre dans lequel les alternatives étaient présentées. Dans cette tâche, où l'appariement des énoncés d'une même paire à l'une ou à l'autre possibilité était par conséquent "co-déterminé", le jugement correct l'emportait dans 90.4%. Au niveau des durées segmentales, les allophones de début de mot étaient plus longs que les allophones de milieu de mot (e.g., "two lips" vs. "tulips") ou de fin de mot (e.g., "new deal" vs. "nude), et les voyelles finales étaient souvent plus longues en position limite qu'en position non-limite (e.g., "grey day" que dans "grade A"). Les deux alternatives se distinguaient également par l'aspiration systématique des plosives sourdes à

[7] Le dièse indique une frontière de mots.

l'initiale de syllabe; des différences spectrales sur /l/ et /r/ entre les positions initiale et finale; la réalisation de /t/ sous sa forme *flap* en fin de syllabe accentuée; la glottalisation des voyelles non-réduites à l'initiale de mot; et un second formant de la voyelle initiale démarrant plus bas.

O'Connor et Tooley (1964) se sont demandés dans quelle mesure le test "perceptif" utilisé par Lehiste (1960) ne surestimait pas la magnitude des variations observées, leur fréquence d'occurrence et leur degré d'implication dans la segmentation. En effet, dans l'étude de Lehiste, les alternatives à prononcer étaient présentées par paires au locuteur, ce qui pouvait l'inciter à maximiser les différences. En outre, la tâche permettait aux auditeurs de comparer les deux alternatives, et donc de baser leur réponse sur les différences entre les deux énoncés plutôt que sur leurs propriétés intrinsèques. C'est pourquoi O'Connor et Tooley ont sélectionné des paires d'énoncés du type "great ape" vs. "grey tape", et les ont faits produire en contexte de phrase par deux locuteurs naïfs. Ici, la tâche perceptive était de choisir entre les deux segmentations possibles celle qui correspondait à l'unique énoncé présenté auditivement. Les auteurs ont trouvé qu'après exclusion des stimuli comportant une plosive glottale devant voyelle à l'initiale de mot, la segmentation correcte n'était perçue que dans 66% des cas. En outre, les énoncés /VC#V/ produisaient 2.4 fois plus d'erreurs que les énoncés /VC#V/. O'Connor et Tooley en ont conclu qu'en situation plus naturelle, le marquage phonétique des frontières lexicales devait être d'une efficacité toute relative.

Au moyen de la parole synthétique, Christie (1974) a examiné si l'aspiration des plosives sourdes était perçue comme un indice de début de mot par les anglophones (cf. Kahn, 1976; Lehiste, 1960). Elle comparait des séquences telles que /eɪstɑː/ vs. /eɪst^{h}ɑː/, qui correspondaient potentiellement à "a star" et "ace tar". A l'aide d'une tâche de catégorisation comparable à celle qu'utilisaient O'Connor et Tooley (1964), elle a trouvé que les choix de segmentation étaient très fortement liés à l'occurrence de la variante aspirée.

Dans un style assez différent, Nakatani et Dukes (1977) ont utilisé de la parole "recombinée" afin de localiser dans le signal les indices de frontières les plus puissants. Au moyen d'hybrides générés en combinant des portions de l'un ou l'autre membre d'une paire d'énoncés homophoniques (e.g., "stay dill" vs. "stayed ill"), ils ont évalué le poids

informationnel de quatre loci: le début de l'énoncé (e.g., "stay"), la fin de l'énoncé (e.g., "ill"), ainsi que les transitions de la première voyelle à la consonne pivotale (ou dyade finale; e.g., "ayed") et de la consonne pivotale à la seconde voyelle (ou dyade initiale; e.g., "di"). Les participants devaient choisir, parmi quatre possibilités (e.g., "stay dill", "stayed ill", "stayed dill" ou "stay ill"), celle qui correspondait le mieux à l'hybride entendu. Les auteurs ont trouvé que les choix de segmentation étaient uniquement influencés par la dyade initiale (en gros, le début du second mot), sauf dans le cas de /l/ et /r/, où la dyade finale (en gros, la fin du premier mot) jouait également un rôle. D'ailleurs, lorsque les dyades finale et initiale provenaient de sources différentes, la consonne pivotale /l/ ou /r/ était perçue comme une géminée ou n'était plus perçue du tout. Des analyses spectrographiques ont révélé que les voyelles de débuts de mots étaient systématiquement précédées d'une plosive glottale, que les consonnes initiales étaient marquées par l'explosion des obstruantes et l'aspiration des plosives sourdes, et que la voyelle du premier mot était plus longue en position limite qu'en position non-limite. Comme seule la dyade initiale (le début du second mot) guidait les choix de segmentations, les auteurs ont conclu que les variations de durée ne jouaient qu'un rôle mineur dans la segmentation, et que l'information essentielle résiderait dans les variations allophoniques (qualitatives) présentes sur les débuts de mots (pour une expérience et des résultats similaires en néerlandais, voir Cohen 1987). Cependant, vu la méthode utilisée par Nakatani et Dukes, il se pourrait que les variations de durée présentes dans les énoncés d'origine aient été rendues inexploitables. Par ailleurs, les études menées par Barry (1981) ont montré que les indices allophoniques comme ceux mis en évidence ici sont nettement plus variables en anglais continu; en particulier, la glottalisation des voyelles de débuts de mots est un phénomène assez rare. Dans l'étude de Nakatani et Dukes, les énoncés de types VC#V (e.g., "known ocean") se caractérisaient donc vraisemblablement par des indices allophoniques très puissants mais peu réalistes.

Jusqu'à présent, la donnée empirique la plus remarquable en faveur de l'utilisation d'indices de durée dans la localisation des frontières de mots a été fournie, pour le néerlandais, par Quené (1992; voir aussi Quené, 1989). Comparant des séquences telles que "die pin" vs. "diep in" dans une tâche où les participants devaient choisir entre deux alternatives celle qui

correspondait le mieux à ce qu'ils entendaient, l'auteur a trouvé que la segmentation choisie était la bonne dans 80.4% des cas. Les énoncés de type V_1#CV_2 se distinguaient de leur homologues essentiellement par trois paramètres de durée: (1) un temps de retombée de l'énergie de la première voyelle V_1 plus long, (2) une consonne pivotale plus longue, et (3) un temps de montée de l'énergie de la seconde voyelle plus court[8]. En outre, des corrélations suggéraient l'existence d'un lien entre les variations observées sur ces trois paramètres et la segmentation perçue par l'auditeur. Afin de déterminer la pertinence de ces paramètres pour la segmentation, l'auteur a dès lors manipulé orthogonalement la durée de la consonne pivotale et le temps de montée en énergie de la seconde voyelle, c'est-à-dire les deux paramètres où les variations étaient les plus fortes. Ceux-ci étaient soit allongés, soit rétrécis de 33%. A l'aide de la même tâche de choix forcé, il a pu établir que les deux paramètres manipulés contribuaient fortement, et indépendamment, à la segmentation. L'occurrence d'un biais sur les séquences V_1C#V_2 en faveur de la segmentation d'origine indiquait néanmoins l'utilisation (et donc la présence) d'autres indices, sans doute qualitatifs. Dans une étude complémentaire, Quené (1993) a reproduit ces résultats en les étendant à des frontières impliquant des suites de consonnes (i.e., VCC(#)V vs. V(#)CCV et VC(#)CV vs. V(#)CCV) et à des énoncés constitués de nonmots, la performance étant la meilleure lorsqu'un accent d'emphase tombait sur le second mot (ou nonmot). Etant donné la différence de classe grammaticale entre les deux alternatives dans le cas des mots (e.g., "die pin", article démonstratif + nom, vs. "diep in", adjectif/ adverbe + préposition), l'obtention de résultats similaires avec des nonmots permet de couper court à l'idée que les effets rapportés dans Quené (1992) ne soient que le reflet de différences prosodiques liées à la structure syntaxique plutôt qu'à la position de la frontière lexicale.

2.1.1.3 Etudes perceptives utilisant une tâche de segmentation *online*

A notre connaissance, seules deux études ont testé l'existence d'indices phonétiques de l'assignation lexicale d'un segment par une tâche de segmentation *online*. Toutes deux

[8] Le temps de décroissance de l'énergie est défini ici comme l'intervalle temporel entre la période d'amplitude maximale et la fin du segment; par temps de montée en énergie, il faut entendre l'inverse.

concernent le phénomène de liaison (enchaînée) en français.[9] A l'aide d'une tâche de détection de phonèmes, Wauquier-Gravelines (1996, expériences 1-4) a montré que l'accès à la consonne intervocalique est plus difficile dans les cas de liaison effective, où cette consonne est flottante (e.g., "un auteur"; "élégant iranien"), que dans les cas de liaison potentielle, où elle constitue l'initiale du mot suivant et a la même identité que le segment latent (e.g., "un notaire"; "élégant tirailleur"; cf. Matter, 1986). Bien que ce résultat soit compatible avec l'idée que le segment de liaison soit perceptivement différent de sa variante initiale, laquelle était d'ailleurs généralement plus longue, il n'en fournit cependant pas une démonstration univoque. Alternativement, on peut expliquer ce résultat en supposant que la différence de performance reflète la différence de statut phonologique du phonème à détecter dans les représentations (lexicales) sous-jacentes (i.e., latent vs. toujours réalisé), et donc d'une certaine opacité du segment flottant.

Une démonstration convaincante provient d'une étude réalisée par Spinelli, McQueen et Cutler (2003) à l'aide du paradigme d'amorçage par répétition intermodale. Les auteurs comparaient l'activation lexicale dans des paires minimales d'énoncés dont l'un était un cas de liaison effective (e.g., "dernier oignon") et l'autre un cas de liaison potentielle (e.g., "dernier rognon"; cf. ci-dessus). Les cibles visuelles étaient précédées par quatre types d'amorces auditives, dont deux étaient ambiguës du point de vue lexical. Dans l'une de ces dernières, la cible (e.g., "oignon") avait été prononcée intentionnellement, alors que dans l'autre (e.g., "dernier rognon"), elle n'apparaissait que de manière fortuite et ne correspondait pas aux intentions lexicales du locuteur. Le même raisonnement s'appliquait pour les séquences non-ambiguës (e.g., "demi-oignon" vs. "demi-rognon"). L'amorçage était testé en décision lexicale, pour le mot débutant par la voyelle comme pour celui débutant par la consonne. Avec les séquences non-ambiguës, Spinelli *et al.* n'ont trouvé de facilitation que lorsque la cible avait été prononcée intentionnellement; en particulier, il n'y avait aucune trace de l'activation du mot commençant par la voyelle en condition non-alignée (e.g., "demi-rognon"). Plus important, les séquences ambiguës produisaient un effet facilitateur sur les deux types de

[9] Selon la plupart des descriptions linguistiques, la liaison enchaînée consiste en ce qu'une consonne latente de fin de mot fasse surface à l'attaque de la syllabe suivante lorsque le mot qui suit commence par une voyelle. Par exemple, "petit avion" est toujours prononcé /pə.ti.ta.vjɔ̃/ (Encrevé, 1988).

cibles, mais dans une mesure beaucoup plus importante lorsque ces dernières se trouvaient dans les intentions lexicales du locuteur (e.g., "dernier oignon") que lorsqu'elles ne s'y trouvaient pas (e.g., "dernier rognon"). Corroborant ces résultats, des analyses acoustiques sur les stimuli et un corpus produit par dix locuteurs naïfs ont montré que la consonne de début de mot était plus longue que celle de liaison (de 20% et 22% pour les deux locuteurs expérimentaux, et de 11% pour les dix locuteurs naïfs).

2.1.2 Attribution de deux syllabes au même mot ou à des mots différents

2.1.2.1 Aperçu des variations phonétiques

Les alternatives de paires minimales telles que "car fourgué" vs. "carrefour gai" diffèrent avant tout par le fait que les voyelles finales sont généralement plus longues que leurs homologues en syllabe non finale. Ce phénomène est attesté notamment en français (Christophe, Dupoux, Bertoncini, & Mehler, 1994; Delattre, 1966; Fougeron *et al.*, 2002; Schaegis, Spinelli, & Welby, 2005), en anglais (Davis *et al.*, 2002; Klatt, 1976; Lehiste, 1972; Mills, 1980; Nakatani & Schaffer, 1978; Umeda, 1975) et en néerlandais (Nooteboom, 1972; Salverda, Dahan, & McQueen, 2004). Des différences au niveau des fréquences formantiques ont également été mises en évidence (voir Schaegis *et al.*, 2005). Comme déjà mentionné plus haut, hormis certaines variantes allophoniques spécifiques des débuts de mots (e.g., la glottalisation des voyelles en anglais), il ne semble pas y avoir de différence systématique et fiable dans l'allongement et dans le renforcement articulatoire entre les allophones à l'initiale de mot et ceux à l'initiale de syllabe en position médiane dans le mot (Krakow, 1989; Fougeron *et al.*, 2002; Keating *et al.*, 1999; Schaegis *et al.*, 2005).

2.1.2.2 Etudes perceptives basées sur les choix de segmentation explicites

Une des premières études perceptives basées sur les choix de segmentation explicites est celle Nakatani et Schaffer (1978), utilisant de la parole réitérée. Cette technique, appliquée ici à l'anglais, consiste à faire produire par le locuteur la même syllabe (e.g., "ma") pour chacune des syllabes d'un énoncé, et permet donc de neutraliser la variabilité liée à la durée intrinsèque des phonèmes originaux, tout en préservant les caractéristiques prosodiques de l'énoncé

(Liberman & Streeter, 1978). Les modèles d'énoncés testés comprenaient un adjectif suivi d'un nom, un des items étant monosyllabique et l'autre bisyllabique, et les séquences de syllabes correspondantes (e.g., "ma#mama" vs. "mama#ma") étaient produites en contexte de phrases. En moyenne, les stimuli obtenus montraient un allongement de la voyelle finale du premier mot (de 25%) et de la consonne initiale du second (de 18%). La tâche des participants étaient de choisir, parmi les deux segmentations possibles (e.g., "bold design" vs. "noisy dog"), celle qui correspondait le mieux au trisyllabe présenté trois fois auditivement. La performance obtenue variait fortement en fonction de la structure accentuelle (i.e., entre 58% et 87%). La meilleure performance était obtenue pour les énoncés impliquant deux accents primaires côte à côte, puisque, dans ce cas, la frontière de mots se situe nécessairement entre les deux syllabes accentuées —une syllabe inaccentuée ne peut être à elle seule ni un adjectif ni un nom. Néanmoins, la performance était supérieure au niveau du hasard également pour les séquences dont la segmentation ne pouvait être déduite de la structure accentuelle. Ainsi, bien qu'un patron "accentuée-inaccentuée-accentuée" corresponde autant à un énoncé tel que "bold design" qu'à un autre tel que "noisy dog", les participants étaient capables de déterminer si les séquences "mamama" correspondantes comportaient une frontière lexicale après la première syllabe ou après la seconde. Les auteurs ont ensuite synthétisé des "mamama" hybrides à structure accentuelle ambiguë, tout en variant, de manière orthogonale, la durée segmentale, l'amplitude, la fréquence fondamentale et les informations spectrales (c'est-à-dire la répartition de l'énergie selon les bandes de fréquence). Ils ont trouvé que la durée segmentale constituait l'information la plus cruciale pour la segmentation de ce type de séquences; lorsque les variations de durée étaient anéanties, la performance de segmentation devenait aléatoire.

Une autre étude est celle menée par Cutler et Butterfield (1990) sur les groupes clitiques de l'anglais. Les auteurs comparaient des paires minimales d'énoncés de deux syllabes produits en contexte de phrase, dont l'un était un groupe clitique, et l'autre, un mot de contenu (e.g., "let us" vs. "lettuce"). Les stimuli se caractérisaient par un allongement des syllabes de fins de mots, de l'ordre de 7.6% pour les syllabes inaccentuées et de 2% pour les syllabes accentuées. Les participants devaient déterminer si le bisyllabe présenté auditivement était constitué d'un

seul mot ou de deux mots, les deux interprétations possibles étant fournies par écrit. Les auteurs ont observé que les participants tendaient fortement à percevoir les énoncés comme étant formés d'un seul mot, avec 70% de réponses "un" pour les bisyllabes, et 63.6% pour les groupes clitiques. Si la différence entre ces deux conditions indique que les indices de frontières lexicales subsistent même à l'intérieur du groupe clitique, la faible performance des participants dans cette condition fait douter de l'exploitabilité de ces indices. Les autres études ayant regardé à la perception de ce type d'ambiguïtés ont utilisé des tâches de segmentation et de reconnaissance des mots en temps réel.

2.1.2.3 Etudes perceptives utilisant une tâche de segmentation *online*

Au moyen du paradigme de détection de phonèmes, Christophe (1993, Expériences 8-9) a examiné l'activation lexicale générée par des paires de séquences momentanément ambiguës, telles que "chat grincheux" vs. "chagrin fou". Dans chaque paire, une des séquences était composée d'un nom monosyllabique et d'un adjectif bisyllabique; inversement, l'autre séquence était composée d'un nom bisyllabique et d'un adjectif monosyllabique. Si la frontière lexicale entre le nom et l'adjectif était signalée par des indices phonétiques, le support lexical à la détection d'un phonème de la première syllabe allait être obtenu plus rapidement lorsque le monosyllabe figurait dans les intentions lexicales du locuteur, que lorsqu'il n'y figurait pas. En accord avec cette hypothèse, l'attaque et la coda de la première syllabe étaient détectées plus rapidement lorsque le nom était monosyllabique. Sur le plan acoustique, l'attaque, la voyelle et la coda de la première syllabe étaient d'ailleurs plus longues pour les séquences débutant par un monosyllabe. Cependant, on peut expliquer ces données autrement: puisque les séquences auditives n'étaient pas complètement ambiguës du point de vue lexical, il est possible que l'activation du premier monosyllabe ait été déterminée par l'assignation lexicale de la dernière syllabe (e.g., "cheux" vs. "fou").

Les autres études échappent à cette critique. Gow et Gordon (1995) ont utilisé l'amorçage sémantique intermodal afin de comparer l'activation lexicale générée par des paires d'énoncés, dont un était constitué d'un mot long, et l'autre d'une séquence homophonique (sur le plan phonémique) composée de deux mots courts (e.g., "tulips" vs. "two lips"). Les deux énoncés

servaient d'amorces, enchâssés dans une même phrase porteuse neutre du point de vue contextuel (e.g., "She placed her tulips/two lips on his cheek."). La tâche des participants était de décider du statut lexical d'une cible visuelle présentée au terme de l'amorce, parfois liée sémantiquement soit au mot long soit au second mot court. Les auteurs ont trouvé que pour les cibles liées au mot long (e.g., "flower"), la séquence de mots courts produisait un effet de facilitation similaire à celui induit par le mot long. En revanche, lorsque la cible était liée au mot court (e.g., "kiss"), le mot long (e.g., "tulips") ne produisait aucun effet, alors que le mot court accélérait la performance aussi bien dans les énoncés ambigus (e.g., "two lips") que dans les énoncés non-ambigus (e.g., "warm lips"). Des analyses de durée ont révélé que l'allophone à l'initiale du second mot était plus long (de 15.7%) que son homologue de milieu de mot. Ces résultats montrent que, en somme, si les auditeurs accèdent, dans certains cas, à des mots qui ne reflètent pas les intentions lexicales du locuteur, des indices acoustico-phonétiques liés aux débuts de mots semblent néanmoins contraindre les processus de segmentation et de reconnaissance.

Davis *et al.* (2002) ont examiné le degré d'activation d'un monosyllabe (e.g., "cap") et d'un bisyllabe (e.g., "captain") dont la syllabe initiale correspondait au monosyllabe, alors que les participants étaient soumis à une phrase momentanément ambiguë dans laquelle était intentionnellement articulé soit le mot court, soit le mot long (e.g., "The soldier saluted the flag with his cap tucked under his arm." vs. "The soldier saluted the flag with his captain looking on."). En *gating* comme en amorçage par répétition intermodale, les auteurs ont trouvé que l'activation de chacun des mots était plus importante lorsque le mot était présent dans les intentions du locuteur que lorsqu'il ne l'était. En outre, dans le second paradigme, le monosyllabe amorçait aussi bien le mot long que le mot court, alors que la première syllabe du bisyllabe n'amorçait que le mot long. Sur le plan acoustique, la voyelle de la séquence ambiguë était plus longue et d'une F0 plus basse lorsque la séquence chevauchait deux mots (e.g., cap tucked). Les productions des mêmes énoncés par trois locuteurs naïfs ont confirmé les différences de durée et de F0, et mis en évidence un allongement de l'attaque de la syllabe suivante lorsque celle-ci était à l'initiale de mot (e.g., cap tucked). Les auteurs ont interprété

ces résultats en termes de détails infra-phonémiques codant au niveau lexical le nombre de syllabes contenues dans le mot.

Dans une étude sur le néerlandais, Salverda *et al.* (2004) ont réexaminé cette hypothèse. A l'aide de l'enregistrement des mouvements oculaires lors de la tâche de sélection d'objets imagés, ils ont comparé l'activation d'un monosyllabe (e.g., "ham") à celle d'un bisyllabe (e.g., "hamster") dans lequel le monosyllabe il était enchâssé à l'initiale; le bisyllabe faisait partie de la phrase présentée auditivement aux participants (e.g., "Zij dacht dat die hamster verdwenen was."). En outre, la première syllabe du bisyllabe provenait soit d'un exemplaire acoustique du monosyllabe (e.g., "ham" extrait de "Zij dacht dat die ham stuckgesneden was."), soit d'un autre exemplaire du bisyllabe. Les auteurs ont trouvé que la probabilité de fixer l'image du compétiteur (e.g., "ham") au détriment de l'image du bisyllabe dépendait de la durée de la portion importée plutôt que de son origine lexicale. Lorsque les monosyllabes importés étaient plus longs —ce qui est généralement le cas (ici de 6.9%)— le taux de fixations du compétiteur était plus élevé avec les monosyllabes qu'avec les syllabes initiales. Par contre, cette différence disparaissait avec des monosyllabes et des syllabes initiales de même durée, et s'inversait lorsque les syllabes initiales étaient plus longues. La perception d'une syllabe comme mot semble donc régie, avant tout, par la durée de la syllabe. Cela suggère que les différences acoustiques entre mots courts et mots longs ne sont pas des propriétés des représentations lexicales (cf. Davis *et al.*, 2002). Pour Salverda *et al.*, les différences observées seraient exclusivement liées à la structure prosodique, spécifiée au niveau post-lexical, et au degré de réalisation des frontières de mots dans cette structure (cf. Nakatani & Schaffer, 1978).

Enfin dans une série d'expériences récentes, Schaegis *et al.* (2005) ont étudié la distinction entre le multisyllabe résultant du processus d'élision en français (e.g., "l'affiche") et une séquence homophonémique dont la première syllabe constitue elle-même un mot (e.g., "la fiche"). Des analyses acoustiques assez exhaustives ont mis en évidence des différences à la fois temporelles et spectrales entre les alternatives: 1) l'article élidé —toujours "l"— était plus long que la consonne correspondante dans l'article non-élidé; 2) la première voyelle, l'attaque de la seconde syllabe et la voyelle de cette syllabe étaient plus courtes dans l'élision que dans

la séquence sans élision —le raccourcissement de la voyelle à l'initiale de mot n'absorbant pas complètement l'allongement de l'article élidé (e.g., "l'affiche"); 3) la première voyelle —toujours "a"— se caractérisait par une F2 plus basse, i.e., position linguale plus antérieure, lorsqu'elle suivait l'article élidé; 4) dans la séquence avec élision toujours, la seconde voyelle avait une F1 plus basse lorsqu'elle était ouverte (e.g., "l'amarre" vs. "la mare"); elle avait aussi une F2 plus basse, i.e., position linguale moins antérieure, lorsqu'elle était antérieure (e.g., "l'amarre" vs. "la mare"), et une F2 plus élevée, i.e., position linguale moins postérieure, lorsqu'elle était postérieure (e.g., "l'allocation" vs. "la location"); enfin 5) le début du mot de contenu étant marqué par une montée de F0, la voyelle à initiale de mot était d'une F0 plus haute que son homologue en finale de l'article non-élidé. Sur le plan perceptif, un paradigme ABX et une tâche de catégorisation ont montré que les participants pouvaient discriminer les alternatives et retrouver la segmentation lexicale voulue par le locuteur, mais la performance dans ces deux tâches était loin d'être optimale (i.e., 68% et 75%, respectivement). De manière convergente, un paradigme d'amorçage par répétition intermodale sur les deux mots de contenu (e.g., "affiche" vs. "fiche") a montré que la facilitation produite en décision lexicale était plus importante lorsque la cible était présente dans les intentions lexicales du locuteur (cf. Spinelli *et al.*, 2003). Autrement dit, si le système perceptif est capable de distinguer les deux alternatives durant la segmentation *online*, leurs différences acoustiques ne semblent pas être assez fortes pour prévenir l'activation du compétiteur, qu'il soit enchâssé ou enchâssant.

2.1.3 En synthèse

Comme on peut le voir, les variations allophoniques induites ou corrélées avec les frontières lexicales sont exploitées au cours de la segmentation. En ce qui concerne les aspects de production, l'occurrence d'une frontière lexicale se marque généralement par une tendance à l'allongement de la voyelle qui précède la frontière, tandis que les débuts de mots ou de syllabes se caractérisent par un allongement et par un renforcement articulatoire qui tendent à distinguer ces segments des allophones non-initiaux. En outre, l'absence d'une différence fiable dans l'implémentation articulatoire des débuts de mots par rapport à celui des débuts de

syllabes suggère que le statut acoustique/phonétique particulier des débuts de mots soit essentiellement lié à leur position d'attaques de syllabes.

En ce qui concerne les aspects de perception, les performances moyennes en tâche de catégorisation suggèrent que les indices infra-segmentaux ne sont pas toujours présents ou, du moins, qu'ils ne sont pas exploités de manière optimale (Barry, 1981; Cutler & Butterfield, 1990; O'Connor & Tooley, 1964; Quené, 1992). De la même façon, les expériences d'amorçage intermodal (Davis *et al.* 2002; Gow & Gordon, 1995; Schaegis *et al.*, 2005; Spinelli *et al.*, 2003) démontrent que ces indices guident les recherches lexicales, mais qu'ils n'empêchent pas l'évocation des alternatives. Les études comparant les suites de monosyllabes à leurs homophones bisyllabiques montrent d'ailleurs une asymétrie intéressante dans le cas de l'anglais, mais pas du français: alors que les monosyllabes (e.g., "cap#t") activent aussi bien les mots courts (e.g., "cap") que les mots longs (e.g., "captain"), les bisyllabes n'activent comparativement que les mots longs (e.g., Davis *et al.*, 2002; Gow & Gordon, 1995).

Enfin, sur le plan méthodologique, trois remarques s'imposent. Premièrement, seules les études de Nakatani & Schaffer (1978) et Quené (1992, 1993) démontrent *stricto sensu* (à l'aide de corrélations) l'existence d'un lien entre les variations acoustiques mises en évidence dans les stimuli et la performance de segmentation des auditeurs. Deuxièmement, à l'exception des observations de Spinelli *et al.* (2003), les données indiquant l'exploitation d'indices infra-phonémiques liés à l'assignation lexicale d'un segment (e.g., V#C vs. VC#) ont été recueillies à l'aide de tâches *offline*, telles que la catégorisation, lesquelles impliquent sans doute plus que la segmentation et la reconnaissance des mots. Troisièmement, excepté Wauquier-Gravelines (1996), dont les résultats restent équivoques, les études sur l'assignation lexicale d'un segment (Cohen, 1987; Nakatani & Dukes, 1977; O'Connor & Tooley, 1964; Quené, 1992, 1993; Spinelli *et al.*, 2003) sont basées exclusivement sur des séquences où la segmentation alternative était fort probablement évidente même pour le locuteur naïf. Par exemple, "dernier rognon" est d'un point de vue phonémique une alternative parfaite à "dernier oignon". Il est donc possible que les locuteurs participant à ces expériences aient cherché à éviter l'ambiguïté dans l'intérêt de l'auditeur par un marquage stratégique des

frontières de mots. Les expériences présentées aux chapitres 4-6 auront pour particularité d'utiliser une méthodologie qui permet d'éviter cette possibilité.

2.2 Indices syllabiques

L'hypothèse que la structure syllabique joue un rôle dans la segmentation est motivée avant tout par le fait que les débuts de mots sont généralement alignés sur les débuts de syllabes. Sous cet angle, les attaques syllabiques pourrait donc offrir des points d'ancrage privilégiés pour le déclenchement des recherches lexicales. C'est exactement ce qu'ont avancé Content, Kearns et Frauenfelder (2001; voir aussi Frauenfelder & Content, 1999). A l'aide de tâches de syllabation explicite, Content *et al.* ont observé que la localisation des débuts de syllabes est plus robuste et moins labile que la localisation des fins de syllabes. Sur base de cette dissociation, ces auteurs ont postulé que le mécanisme sous-tendant le repérage des attaques syllabiques pourrait être à la base d'une heuristique de segmentation prélexicale (*Syllable Onset Segmentation Heuristic*; SOSH), dans laquelle chaque début de syllabe est considéré par l'auditeur comme étant un début de mot potentiel.

L'hypothèse de segmentation aux attaques syllabiques est étayée par plusieurs résultats dans des tâches impliquant la reconnaissance des mots. Dans une expérience d'amorçage intermodal en néerlandais, Vroomen et de Gelder (1997a) ont examiné l'activation d'un monosyllabe enchâssé en finale d'un mot plus long en mesurant le taux de facilitation sur un mot visuel lié sémantiquement à l'enchâssé. Ils ont mis en évidence que les mots alignés sur un début de syllabe étaient activés (e.g., "boos" dans "fram<u>boos</u>"), mais n'ont trouvé aucune trace de l'activation de mots non-alignés avec le début d'un monosyllabe (e.g., "wijn" dans "z<u>wijn</u>").

Dans une étude de *word spotting*, Dumay, Frauenfelder et Content (2002)[10] ont évalué le rôle des débuts et des fins de syllabes dans la segmentation lexicale en français. Les participants devaient détecter des mots de structure CVC (e.g., "lac") enchâssés à l'initiale ou à la finale de bisyllabes sans signification. Dans la situation d'enchâssement final (Expérience 1), le début des cibles était soit aligné soit non-aligné avec le début de la seconde syllabe (e.g.,

[10] L'article de Dumay *et al.* (2002) paru dans *Brain and Language* est reproduit intégralement à l'Appendice 1.

"zun.lac" vs. "zu.glac"). Inversement, dans la situation d'enchâssement initial (Expériences 2-3), la fin des cibles était soit alignée soit non-alignée avec la fin de la première syllabe (e.g., "lac.tuf" vs. "la.cluf"). En accord avec l'idée que seul l'alignement entre le début du mot et le début d'une syllabe est crucial pour l'accès lexical, le coût perceptif induit par le non-alignement initial était plus important et plus robuste (sur le plan statistique) que celui induit par le non-alignement final. Notons que cette asymétrie est difficilement conciliable avec l'hypothèse avancée par Mehler et coll. (Mehler, 1981; Mehler *et al.*, 1990; Seguí, 1984; Seguí *et al.*, 1990), selon laquelle, en français et dans les autres langues romanes, la syllabe servirait d'unité de classification en vue de l'accès au lexique. En effet, sur base de cette hypothèse, on s'attendrait à obtenir des effets d'alignement syllabique initial et final équivalents.

Des résultats similaires ont été obtenus par McQueen (1998) en néerlandais. Dans cette étude, la détection de mots enchâssés en position finale était plus facile lorsque le début des cibles était aligné avec la structure syllabique (e.g., "rok" dans "fim.rok" vs. "fi.drok") que lorsqu'il ne l'était pas, et un effet d'alignement était aussi obtenu pour les cibles en position initiale (e.g., "vel" dans "vel.brul" vs. "velm.brul"). Toutefois, sur les erreurs comme sur les latences de détection, l'effet d'alignement syllabique était plus marqué pour les débuts que pour les fins de syllabes.

Enfin, en anglais et toujours à l'aide de la tâche de *word spotting*, Weber (2000) a reproduit l'effet de non-alignement initial sur des monolingues et des bilingues allemand-anglais (e.g., "mar.fluck" vs. "pun.luck"). Dans leur ensemble, ces études suggèrent donc que l'alignement sur les attaques de syllabes guide les recherches lexicales non seulement en français, où la structure syllabique est relativement claire, mais aussi dans les germaniques, où elle est plus opaque.

2.3 Indices phonotactiques et probabilités transitionnelles

2.3.1 Restrictions phonotactiques

Une source d'information potentiellement pertinente pour la segmentation provient des propriétés phonotactiques propres à chaque langue (Church, 1987; Harrington, Watson, & Cooper, 1989; Gimson, 1980; Harris, 1955). La phonotactique fait référence aux contraintes sur l'ordre des segments phonémiques à l'intérieur des syllabes, morphèmes et mots d'une langue donnée (Trask, 1996). Cette dimension est généralement décrite par un ensemble de restrictions sur les co-occurrences possibles à l'intérieur de ces unités. Si elles reflètent dans une large mesure la dynamique articulatoire de la langue, ces restrictions séquentielles apportent une information indépendante du mode de réalisation des segments. Par exemple, en français, la séquence /ml/ n'est pas permise à l'intérieur d'une syllabe, la séquence /voyelle nasale + m/ n'apparaît jamais à l'intérieur d'un morphème, et les séquences de plus de quatre consonnes telle que /ktstr/ n'existent dans aucun mot. Un francophone rencontrant de telles séquences peut être sûr qu'il s'agit respectivement de frontières syllabique (e.g., "camelote"), morphémique (e.g., "emmener") et lexicale (e.g., "pacte stricte"), et exploiter cette information pour contraindre l'accès au lexique.

Au niveau empirique, les études sur l'effet d'alignement syllabique présentées dans la section précédente (Dumay *et al.*, 2002; McQueen, 1998; Vroomen & de Gelder, 1997; Weber, 2000) ont toutes eu recours aux restrictions phonotactiques pour induire chez le locuteur telle ou telle syllabation des stimuli (e.g., "zun.lac" vs. "zu.glac"). Les effets perceptifs obtenus dans ces études pourraient donc en partie refléter l'utilisation de connaissances sur ces restrictions au cours de la segmentation.

Toutefois, il existe des démonstrations plus convaincantes. Dans son étude de *word spotting* sur les bilingues allemand-anglais, Weber (2000) a montré que la phonotactique de la langue native influence la segmentation de la seconde langue. Les participants détectaient plus facilement les cibles non-alignées sur la structure syllabique de l'anglais, telles que "weapon" dans "moy.tweapon" lorsque la phonotactique de l'allemand favorisait leur alignement (e.g., "moyt.weapon") que lorsqu'elle ne le favorisait pas (e.g., "ru.kweapon").

Autre exemple, toujours à l'aide du *word spotting*, Suomi, McQueen et Cutler (1997) ont testé un cas particulier de restriction phonotactique attestée en finnois: l'harmonie vocalique. Dans cette langue, si un mot (non-composé) comporte une voyelle antérieure (i.e., /y/, /ø/ et /æ/), il ne peut comporter de voyelles postérieures (i.e., /u/, /o/et /a/) et vice versa, alors que les voyelles neutres (i.e., /i/ et /e/) se combinent au niveau lexical aussi bien avec les voyelles antérieures que postérieures (Karlsson, 1983). Tout clash harmonique, par exemple, d'antérieure à postérieure, entre deux voyelles consécutives indique dès lors fortement qu'il doit y avoir une frontière lexicale quelque part entre les deux voyelles. Les finnois exploitent ces discordances au cours de la segmentation *online*. Les mots à voyelles postérieures, tels que "palo" (*feu*), sont détectés plus facilement dans des séquences dysharmoniques, telles que "kypalo", où la voyelle contextuelle est de l'autre classe, antérieure, que dans des séquences harmoniques, telles que "kupalo", où cette voyelle est de la même classe, postérieure; et idem pour les mots à voyelles antérieures (voir Vroomen, Tuomainen, & de Gelder, 1998, pour des résultats convergents). Les restrictions sur les co-occurrences phonémiques jouent donc un rôle dans la segmentation.

2.3.2 Phonotactique probabiliste

Néanmoins, comme le soulignait déjà Harris (1954, 1955), les contraintes phonotactiques ne doivent pas être absolues (i.e., légale vs. illégale), elles peuvent être probabilistes, certaines séquences de phonèmes étant simplement plus fréquentes que d'autres en une position donnée. L'approche probabiliste de la phonotactique est basée sur l'idée que la séquenciation des phonèmes est nécessairement plus contrainte à l'intérieur des unités linguistiques d'un niveau donné qu'entre deux unités du même niveau. Autrement dit, la probabilité liée à l'occurrence d'un phonème est plus forte si les phonèmes qui le précèdent immédiatement font partie de la même unité que s'ils font partie d'une autre unité. Par exemple (tiré de Brent, 1999), en anglais, le phonème /ð/ est la plupart du temps suivi d'une voyelle, comme dans les mots très fréquents "the", "this", "that" et "them", et seule une faible proportion des occurrences sont suivies d'une consonne. Par conséquent, si une consonne, disons /m/, suit /ð/,

comme dans "bathe more", la faible probabilité de la séquence /ðm/ en anglais continu suggère que /m/ est à l'initiale du mot suivant.

Plusieurs modèles computationnels de la segmentation au cours de l'acquisition lexicale (Cairns, Shillcock, Chater, & Levy, 1997; Christiansen, Allen, & Seidenberg, 1998; Elman, 1990) ont démontré l'utilité des probabilités phonotactiques pour localiser les frontières de mots. Par exemple, à l'aide d'un corpus conversationnel ne comportant aucun marqueur acoustique, Cairns *et al.* (1997) ont montré qu'un réseau récurrent simple (SRN) dont la tâche était de prédire le segment phonétique subséquent parvenait à trouver un nombre substantiel de frontières lexicales sur la seule base des probabilités séquentielles entre les segments: là où le réseau a du mal à prédire le segment suivant, il postule une frontière de mots. En outre, plusieurs études ont montré que les bébés dès l'âge de neuf mois sont sensibles à ces probabilités (Friederici & Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk, Luce, & Charles-Luce, 1994) et les utilisent pour découvrir les portions du signal qui correspondent à des mots (Mattys, Jusczyk, Luce, & Morgan, 1999; Mattys & Jusczyk, 2001).

L'exploitation des probabilités phonotactiques dans la segmentation lexicale chez l'adulte est confirmée par plusieurs résultats de *word spotting*. A l'aide du néerlandais, van der Lugt (2001) a examiné si la performance de détection était affectée par la probabilité des diphones situés de part et d'autre de la frontière (syllabique) séparant des cibles initiales ou finales de leurs contextes adjacents (e.g., CVC.CVC vs. CVC.CVC). Il a trouvé que seule la probabilité de la portion CV à l'initiale du mot cible influençait la détection, les cibles à portion fréquente étant détectées plus rapidement que les cibles à portion rare (e.g., "hien.galg" vs. "pien.geur"). Cependant, puisque cette étude ne teste que l'influence des diphones internes, elle ne permet pas de dire si les probabilités transitionnelles peuvent servir de marqueurs de frontières de mots. C'est la question qu'ont examiné Gaygen et Luce (2002) à l'aide de l'anglais. Ces auteurs ont manipulé la probabilité des diphones impliquant le phonème initial ou final de cibles CVC enchâssées en milieu de séquences CVC.CVC.CVC. Les latences de détection étaient plus courtes lorsque le diphone chevauchant la frontière initiale de la cible (i.e., CVC.CVC.CVC) était rare du point de vue phonotactique que lorsqu'il était fréquent, alors qu'un contraste

similaire à la frontière finale (i.e., CVC.CVC.CVC) ne générait aucun effet. Cette asymétrie s'explique par le fait que seul le repérage des débuts de mots est pertinent pour la segmentation, les fins de mots pouvant être déduites sur base des connaissances lexicales. En revanche, pour ce qui est des diphones internes, et en contradiction avec les résultats de van der Lugt (2001), c'est en fin de cible qu'il y avait un effet (i.e., CVC.CVC.CVC), les latences de détection étant plus courtes pour les mots à diphone final fréquent que pour ceux à diphone final rare. Ces données montrent donc que plus un diphone est probable, plus l'auditeur tend à le considérer comme intra-lexical; et plus il est improbable, plus l'auditeur tend à considérer les phonèmes qui le constituent comme appartenant à des mots différents.

2.3.3 Probabilités transitionnelles non-phonémiques

La sensibilité de l'auditeur aux probabilités transitionnelles n'est par ailleurs pas restreinte à celles qui ont le phonème comme primitive (i.e., phonotactiques). Saffran et coll. (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997) ont montré qu'aussi bien des adultes que des bébés de 8 mois confrontés à un langage artificiel présenté sous forme de parole concaténée (e.g., "babupubupadadutabapatubipidabututibu") tendaient à inférer les frontières de mots là où les probabilités transitionnelles entre deux syllabes adjacentes étaient les plus faibles. Après plusieurs minutes d'exposition à ce type de continuum, les trisyllabes délimités par des creux dans les probabilités transitionnelles (e.g., "bupada") étaient jugés comme plus familiers que ceux dont les limites se situaient entre deux syllabes successives hautement probables (e.g., "pubupa"). Notons enfin que ce genre de phénomène a également été démontré pour des manipulations impliquant les probabilités transitionnelles entre syllabes non-adjacentes, telles que "bu" et "da" dans "bupada" (Peña, Bonatti, Nespor, & Mehler, 2002).

Pris dans leur ensemble, ces résultats indiquent donc que des connaissances sur les probabilités d'occurrence des séquences d'unités infralexicales sont utilisées pour localiser les portions susceptibles de correspondre à des mots dans le continuum.

2.4 Indices rythmiques

Dans chaque langue, il existe une plus ou moins grande corrélation entre certains éléments récurrents liés au rythme et la position relative des frontières lexicales. Sur base de l'occurrence de ces éléments, les auditeurs pourraient donc faire des inférences quant à la localisation approximative des débuts (ou des fins) de mots. Afin d'illustrer ce mécanisme, nous examinerons le rôle joué dans la segmentation par trois de ces éléments: la qualité vocalique en anglais, celui de l'accent (ou *stress*) lexical en néerlandais, et celui de l'accent métrique en français.

2.4.1 Qualité vocalique (en anglais)

La notion de qualité vocalique fait référence à cette dimension des voyelles qui permet de distinguer en anglais entre les syllabes fortes, dont les voyelles sont pleines (i.e., non-réduite), et les syllabes faibles, dont les voyelles sont réduites (habituellement en schwa). Dans cette langue, la majorité des mots de contenu débutent par une syllabe forte et la plupart des syllabes fortes sont à l'initiale d'un mot de contenu; en outre, la plupart des syllabes faibles sont à l'initiale d'un mot de fonction (Cutler & Carter, 1987; Shillcock, Bard, & Spensley, 1988). Par conséquent, une heuristique de segmentation qui devrait être efficace pour cette langue serait de prendre chaque syllabe forte comme étant potentiellement le début d'un mot de contenu. C'est précisément ce dont Cutler et Norris (1988) ont fait l'hypothèse au travers de la *Stratégie de Segmentation Métrique* (MSS; voir Cutler & Norris, 2002, pour une revisite récente).

Une démonstration convaincante de l'utilisation de cette heuristique par les anglophones provient d'une étude de Cutler et Butterfield (1992) sur les erreurs de perception, spontanées ou induites en laboratoire.[11] Ces auteurs ont trouvé que lorsque les auditeurs segmentent mal un énoncé, ils tendent à le faire d'une manière qui favorise l'émergence de la structure métrique dominante trochaïque (i.e., fort-faible): ils insèrent les frontières lexicales le plus

[11] En situation d'induction, les participants ont pour tâche d'identifier des énoncés présentés juste au seuil individuel de perception.

souvent devant les syllabes fortes et les supprimant le plus souvent devant les syllabes faibles, comme dans "conDUCT asCENTS upHILL" perçu "the DOCtor SENT her BILL"[12,13].

Le recours à MSS par les anglophones a aussi été confirmée par des données de *word spotting* (Cutler & Norris, 1988; McQueen *et al.*, 1994; Norris *et al.*, 1995), et plus récemment d'amorçage par répétition intermodale (Mattys, 2004; voir aussi Mattys, White, & Melhorn, 2005). Par exemple, à l'aide du *word spotting*, Cutler et Norris (1988) ont trouvé qu'il était plus facile de détecter un mot à consonne finale devant une voyelle réduite que devant une voyelle pleine. Ainsi, un mot tel que "mint" était plus facile à détecter dans "MINtesh" que dans "MINTAYVE". Ceci est totalement conforme à l'heuristique MSS, selon laquelle l'occurrence d'une voyelle pleine devrait induire un point de segmentation au début de la syllabe forte qui la contient, d'où la nécessité de réassembler le matériel linguistique situé de part et d'autre du point de segmentation.[14]

2.4.2 Accent lexical (en néerlandais)

La distinction entre voyelles pleines et voyelles réduites existe aussi en néerlandais, et dans cette langue, la majorité des mots de contenu débutent par une syllabe forte (Vroomen & de Gelder, 1995). Cependant, la voyelle de la plupart des syllabes inaccentuées n'est pas réduite (Quené & Koster, 1998; Quené & Smith, 1992), au contraire de l'anglais (Fear, Cutler, & Butterfield, 1995; voir aussi chapitre 1, section 2.2). En néerlandais, la qualité pleine ou réduite de la voyelle n'est donc pas un bon prédicteur du statut métrique d'une syllabe, et il est par conséquent improbable que l'heuristique MSS y soit d'une grande utilité.

Puisque dans cette langue, la syllabe initiale des mots porte généralement l'accent primaire (Kager, 1989; van der Hulst, 1984), les propriétés rythmiques les plus susceptibles (d'un point de vue statistique) d'indiquer les débuts de mots sont celles par lesquelles l'accent lexical

[12] Les syllabes fortes sont en majuscules, les syllabes faibles, en minuscules.
[13] Vroomen, van Zon et de Gelder (1996) ont obtenu des résultats similaires en néerlandais.
[14] Notons que l'effet observé ici pourrait tout aussi bien refléter des différences acoustiques et/ou perceptives dans la syllabation des séquences porteuses, une alternative jamais vraiment abordée par les auteurs. En raison de la nature accentuée de la première voyelle et réduite, et donc inaccentuée de la seconde, la consonne finale de la cible dans une séquence telle que "MINtesh" appartient vraisemblablement à la fois à la première et à la seconde syllabe, un phénomène appelé "ambisyllabicité" (Kahn, 1976). Sous cet angle, ce serait donc l'alignement partiel de la fin de la cible avec la fin de la première syllabe caractérisant les séquences "MINtesh" mais pas les séquences "MINTAYVE" qui serait à la source de l'effet.

primaire se manifeste, i.e., allongement de la voyelle, intensité plus forte et spectre plus homogène (Quené & Koster, 1998).

L'utilisation par les néerlandophones d'une heuristique de segmentation basée sur les degrés d'accent lexical est confirmée par des données de *word spotting*. Vroomen & de Gelder (1997b) ont montré que les latences de détection de mots trochaïques de deux syllabes, tels que "KRAter" (prononcé ˈkrɑtər), en finale de trisyllabes sans signification, sont plus courtes lorsque la première syllabe du mot est réalisée avec l'accent primaire (e.g., ˌpɔˈkrɑtər[15]) qu'avec l'accent secondaire (e.g., ˈpɔˌkrɑtər). Notons que ce résultat ne peut-être expliqué par MSS, puisque ici toutes les voyelles sont pleines. Par ailleurs, Vroomen *et al.* (1998) ont trouvé que dans une situation d'apprentissage d'un langage artificiel à partir d'un continuum ininterrompu (cf. Saffran, Newport *et al.*, 1996), les mots d'un langage ne contenant que des voyelles pleines étaient mieux reconnus s'ils avaient un accent initial qu'un accent non-initial.

Dans l'ensemble, ces résultats suggèrent donc que l'accent lexical peut avoir une fonction instrumentale dans la segmentation de la parole.

2.4.3 Accent métrique (en français)

En français, l'accent n'a aucune valeur distinctive, mais néanmoins sa position est fixe. Il se situe sur la syllabe finale d'unités rythmiques (i.e., les groupes accentuels) n'excédant généralement pas plus de deux ou trois syllabes (Dell & Vergnaud, 1984; Tranel, 1987; Wenk & Wioland, 1982). De ce fait, la dernière syllabe des mots polysyllabiques est généralement accentuée. Ce phénomène se manifeste essentiellement par un allongement substantiel de la voyelle par rapport aux voyelles précédentes et un mouvement de F0, le plus souvent descendant (Fletcher, 1991; Pasdeloup, 1990; Vaissière, 1983; 1992). La structure métrique bref-long (ou iambe) constitue donc l'unité rythmique de base. Etant donné la position invariable de l'accent dans leur langue, les auditeurs francophones rencontrant une syllabe allongée peuvent en inférer à coup sûr qu'il s'agit d'une fin de mot.

L'exploitation de la durée relative des syllabes comme marqueur des frontières lexicales en français a été mise en évidence par Banel et Bacri (1994, voir Rietveld, 1980, pour une étude

[15] Les symboles ˈ et ˌ indiquent respectivement les accents primaire et secondaire.

préliminaire). Demandant à leurs participants de juger du nombre de mots dans des séquences telles que /fuʀmi/, interprétables comme un seul bisyllabe (e.g., "fourmi") ou comme deux monosyllabes (e.g., "fourmi" vs. "four" et "mi"), ils ont trouvé qu'il y avait moins de segmentations lorsque les séquences présentaient le patron iambique, typique du français, que lorsqu'elles présentaient le patron trochaïque, inhabituel dans cette langue; un résultat reproduit par Meunier, Frauenfelder et Content (1996). A l'aide du *word spotting*, Banel & Bacri (1997) ont en outre montré qu'il était plus facile de détecter un mot monosyllabique (e.g., "lampe", lɑ̃p) enchâssé à l'initiale dans une séquence trochaïque (e.g., lɑ̃ːpzɔk[16]) que dans une séquence iambique (e.g., lɑ̃pzɔːk). Dans la même veine, Banel, Frauenfelder et Perruchet (1998) ont observé que l'apprentissage de mots de trois syllabes à partir d'un continuum ininterrompu était meilleur pour les "mots" à patron accentuel typique, iambique (bref-bref-long), que pour d'autres à patron rythmique trochaïque (long-bref-bref) ou neutre (long-long-long). Notons que les rapports de durée utilisés par Banel et coll. ont été obtenus à partir des syllabes initiale et finale de mots prononcés isolément. Coupant court à l'idée que les différences perceptives observées par ces auteurs soient spécifiques aux mots isolés, Goetry et Kolinsky (2000) ont obtenu des résultats similaires à l'aide d'un paradigme impliquant des phrases "naturelles" et continues.

3 Approches combinées et hypothèse de "Possible Word Constraint"

Les solutions lexicales et pré-lexicales à la segmentation ne sont pas mutuellement exclusives mais complémentaires. Et jusqu'à présent, cette complémentarité a été envisagée de trois manières différentes. Selon une première conception, les indices dérivés du signal renseignent l'auditeur sur les débuts de mots potentiels, autrement dit, sur les points du continuum où les recherches lexicales doivent être déclenchées, l'accès au lexique n'étant pas continu. Puisque le déclenchement des recherches lexicales n'a lieu que si tel ou tel indice est

[16] Le symbole ː indique l'allongement de la voyelle.

présent, cette conception est qualifiée de "déterministe". Elle se retrouve notamment dans les premières formulations de MSS (Cutler & Norris, 1988; Cutler 1990).

Une deuxième conception considère que les indices de segmentation sont utilisés pour pondérer ou contraindre l'activation des compétiteurs lexicaux évoqués en tout point du continuum comme dans TRACE (McClelland & Elman, 1986) et Shortlist (Norris, 1994). Toute chose étant égale par ailleurs, les candidats alignés sur un début de mot potentiel sont plus activés que ceux qui ne sont pas alignés (voir McQueen *et al.*, 1994; Norris *et al.*, 1995, pour des simulations sur Shortlist). D'un point de vue computationnel, ceci permet notamment au processus de compétition lexicale de converger plus rapidement vers une segmentation optimale de l'énoncé, et donc de stocker pour une période moins longue l'information non-segmentée (voir ce chapitre, section 2.2.2). Par ailleurs l'importance accordée par l'auditeur aux différents indices et stratégies lexicales varie fort probablement en fonction de la qualité du signal et de sa nature plus ou moins ambiguë, comme le suggèrent les données récentes de Mattys et coll. (Mattys, 2004; Mattys *et al.*, 2005).

Enfin, une troisième conception est incarnée par l'hypothèse de "Possible Word Constraint" (PWC) avancée par Norris, McQueen, Cutler et Butterfield (1997). Ces auteurs ont trouvé que dans la tâche de *word spotting*, il est plus difficile de détecter un mot (e.g., "apple") lorsque le contexte adjacent est constitué d'une consonne que lorsqu'il comporte une voyelle (e.g., "fapple" vs. "vuffapple"). Sur base de simulations à l'aide de Shortlist, Norris *et al.* ont suggéré que le système de perception devait inhiber tout candidat lexical potentiellement présent dans le continuum si la portion de parole située entre sa limite initiale (ou finale) et une frontière lexicale très probable (comme celle indiquée par la fin du silence) ne pouvait elle-même constituer un mot. Et les auteurs d'ajouter sur base de nouveaux résultats (Cutler, Demuth, & McQueen, 2002; Cutler, McQueen, Norris, & Somejuan, 2001; Norris, McQueen, Cutler, Butterfield, & Kearns, 2001) que la notion de mot dans PWC fait référence à ce qui définit un mot non pas dans une langue donnée mais à travers toutes les langues: sa nature syllabique (ou vocalique). Selon cette théorisation, les indices prélexicaux nourrissent PWC en signalant les frontières lexicales probables, et ce n'est qu'indirectement, à travers cette

heuristique, qu'ils biaisent le processus de compétition lexicale (voir McQueen, Otake, & Cutler, 2001, pour des résultats contradictoires).

Chapitre 3 Esquisse de la contribution expérimentale

1 Arrêt sur image

Comme nous l'avons vu au chapitre précédent, Content et coll. (Content *et al.* 2001; Dumay *et al.*, 2002; Frauenfelder & Content, 1999) ont fait l'hypothèse d'une heuristique de segmentation prélexicale (SOSH) selon laquelle les débuts de syllabes constituent des points d'alignement privilégiés pour le déclenchement des recherches lexicales. Cette hypothèse est motivée par l'examen des choix de syllabation explicite des locuteurs francophones, qui montre que le repérage des débuts de syllabes est nettement plus fiable et moins labile que le repérage des fins de syllabes, les deux opérations faisant vraisemblablement l'objet de mécanismes différents. A l'appui de SOSH, des études d'amorçage sémantique intermodal en néerlandais (Vroomen & de Gelder, 1997a), et de *word spotting* en néerlandais (McQueen, 1998), en anglais (Weber, 2000) et en français (Dumay *et al.*, 2002) démontrent que les mots alignés sur une frontière de syllabes sont évoqués plus facilement que les mots non-alignés de la sorte (e.g., "zun.lac" vs. "zu.glac"). En outre, comme le mettent en évidence les études de McQueen (1998) et Dumay *et al.* (2002), le non-alignement entre débuts de mots et débuts de syllabes induit un coût perceptif nettement plus important et beaucoup plus robuste que le non-alignement entre fins de mots et fins de syllabes (e.g., "lac.tuf" vs. "la.cluf"). Ceci suggère que si la structure syllabique joue un rôle dans la segmentation de la parole, c'est essentiellement en fournissant des points d'alignement lexical et non en délimitant des unités de classification (voir Dumay *et al.*, 2002, pour une discussion).

Cependant, la viabilité de cette heuristique est compromise par l'existence potentielle du phénomène de resyllabation, dans lequel la structure syllabique dite "de surface", i.e., celle qui est effectivement produite, ne correspond pas à celle attendue sur base des représentations lexicales (Anderson, 1982; Clements & Keyser, 1983; Encrevé, 1988; Grammont, 1933; Léon, 1966; Martinet, 1945; Noske,1982; Nyrop, 1924; Passy, 1917; Pulgram, 1970). En français, trois processus phonologiques sont généralement considérés comme induisant une resyllabation de la chaîne parlée: l'*élision*, dans laquelle deux consonnes s'enchaînent du fait de l'effacement d'une voyelle schwa les séparant (comme dans "pas de rôle" prononcé "pas

d'rôle" [pa.dʀol]); l'*enchaînement*, où la consonne finale d'un mot est rattachée à une voyelle, ou dans certains cas, à une consonne, à l'initiale du mot suivant (comme dans "danse errante" [dɑ̃.sɛʀɑ̃t]); et la *liaison enchaînée*, où une consonne latente en finale de mot fait surface pour être liée à la voyelle à l'initiale du mot suivant (comme dans "petit ami" [pə.ti.ta.mi]). Dans chacun de ces cas, débuts de syllabes et débuts de mots ne coïncidant vraisemblablement pas, l'alignement lexical réalisé sur base des débuts de syllabes, tel que le suggère SOSH, devrait être incorrect.

L'idée de resyllabation au travers des frontières de mots se retrouve à l'avant d'un des modèles de production de la parole les plus connus, celui proposé par Levelt (1989, 1992, 1999; Levelt, Roelofs, & Meyer, 1999). Selon ce modèle, les représentations lexicales ne contiendraient aucune information sur la structure syllabique; au contraire, la syllabation serait déterminée "en chemin", tardivement et sans tenir compte des frontières lexicales, au cours de l'encodage phonologique. Ce-dernier implique l'insertion séquentielle des segments de plusieurs formes lexicales dans le squelette métrique de la fenêtre articulatoire, laquelle n'est jamais inférieure au mot prosodique (Wheeldon & Lahiri, 1997, 2002). Une fois l'encodage phonologique terminé, les syllabes par-là constituées serviraient de code d'accès à des patrons articulatoires syllabiques et relativement abstraits, stockés dans un répertoire phonétique ou calculés en temps réel (cf. Crompton, 1982; Levelt & Wheeldon, 1994; voir aussi Cholin, Schiller, & Levelt, 2004). Ne laissant en surface aucune trace des frontières lexicales, un tel processus devrait donc nécessairement engendrer des décalages fréquents entre débuts de syllabes et débuts de mots au niveau phonétique, et partant rendre l'heuristique de segmentation aux attaques syllabiques inappropriée voire perturbatrice.

A l'encontre de la notion de resyllabation, certaines analyses phonétiques du français suggèrent pourtant la persistance en surface de traces des frontières lexicales, notamment sur les durées segmentales (élision du schwa: Fougeron & Steriade, 1997; Rialland, 1986; enchaînement: Content *et al.*, 2004; Fougeron *et al.*, 2002; liaison enchaînée: Dejean de la Bâtie, 1993; Spinelli *et al.*, 2003; Wauquier-Gravelines, 1996). Mais dans cette langue, la question de l'exploitation perceptive de ces différences de nature infra-phonémique n'a pratiquement pas été étudiée. Les seuls résultats non-équivoques sont ceux d'amorçage

intermodal obtenus récemment sur la liaison par Spinelli *et al.* (2003). Dans cette étude, l'influence bénéfique induite par la présentation préalable du mot cible (e.g., "oignon") dans une séquence lexicalement ambiguë était plus importante lorsque le mot figurait dans les intentions lexicales du locuteur que lorqu'il n'y figurait pas (e.g., "dernier oignon" vs. "dernier rognon"; voir chapitre 2, section 2.1.1, pour plus de détails). Notons cependant que si les auteurs concluent que la consonne de liaison (e.g., /ʀ/ dans l'exemple) doit avoir gardé des traces de son assignation lexicale, ils semblent ne pas mettre en doute l'existence du phénomène de resyllabation. En outre, la question de savoir si les différences acoustico-phonétiques observées (dont la durée plus longue de la consonne à l'initiale de mot) sont des indices de frontière lexicale ou de frontière syllabique n'est pas abordée.

La liaison est un phénomène peu attesté à travers les langues (cf. Sapir, 1933), et il ne concerne généralement que quelques phonèmes du répertoire (e.g., en français, /g/, /n/, /p/, /ʀ/, /t/ et /z/). Il est donc difficile de dire si les indices mis en évidence par Spinelli *et al.* (2003) reflètent une propension du système phonologique à marquer toutes les frontières lexicales ou seulement cette idiosyncrasie que représente la liaison.

La présente recherche a examiné la question de la resyllabation en français à partir de la situation plus fréquente de l'*enchaînement* (e.g., "danse errante"). Plus précisément, elle a cherché à déterminer dans ce cas relativement ambigu l'existence de variations acoustiques/phonétiques liées à la position des frontières de mots, et leur exploitation au cours de la segmentation lexicale. Les conséquences perceptives de l'enchaînement, qui sont évaluées généralement par rapport à une situation où la consonne pivotale est à l'initiale du mot suivant, n'ont jusqu'à présent été étudiées qu'à l'aide de tâches *offline*, telles que l'appariement co-déterminé (Lehiste, 1960) et la catégorisation (Barry, 1984; Nakatani & Dukes, 1977; O'Connor & Tooley, 1964; Quené, 1992, 1993; voir chapitre 2, section 2.1.1, pour plus de détails). Par exemple, à l'aide de la catégorisation de séquences du type "diep in" vs. "die pin", Quené (1992) a montré que les néerlandophones exploitent des variations dans la durée de la consonne pivotale et dans le temps de montée en énergie de la voyelle subséquente pour juger de la segmentation entendue. En outre, elles ne semblent jamais avoir

été étudiées en français, où pourtant les variations allophoniques positionnelles ne sont pas aussi évidentes que dans les autres langues considérées.

2 Considérations méthodologiques

Nous nous sommes centrés sur les situations d'enchaînement impliquant un groupe consonantique, autrement dit, où la consonne pivotale est liée à une autre consonne à l'initiale du mot suivant (e.g., "tant' roublarde", potentiellement syllabifié /tɑ̃.tʀu.../). Il y a deux raisons à ce choix. Primo, l'utilisation des groupes consonantiques permet une comparaison directe avec l'étude de Dumay *et al.* (2002) sur le rôle de la structure syllabique dans la segmentation lexicale (e.g., "zun.lac" vs. "zu.glac"). Puisque dans cette étude, l'alignement entre syllabes et mots était régi par la légalité phonotactique du groupe consonantique en début et fin de syllabe, il est difficile de savoir si l'effet d'alignement syllabique obtenu dans la situation d'enchâssement final résulte de l'utilisation *online* par l'auditeur de connaissances sur les restrictions phonotactiques de langue, et/ou de sa sensibilité à la présence d'indices phonétiques de syllabation. Comme elles gardent constante l'identité des consonnes médianes, les comparaisons entre enchaînement (e.g., "tant#rou") et situation de consonne pivotale initiale (e.g., "tan#trou") rapportées dans les chapitres suivants permettront de répondre en partie à cette question.

Secundo, l'utilisation de séquences à groupe consonantique médian rend possible l'étude de la perception de l'enchaînement au moyen du *word spotting*, lequel requiert un nombre substantiel de cibles lexicales. En revanche, l'emploi de cette tâche semble difficile dans le cas d'une consonne pivotale intervocalique (e.g., "cale égale" vs. "cas légal"). En effet, la langue française ne compte qu'un nombre restreint de mots débutant par une seule consonne et enchâssant à partir de leur second phonème un mot de la même catégorie grammaticale (tels que "légal" enchâssant "égal"), et par-là susceptibles d'être utilisés comme second constituant. Jusqu'à présent, la seule situation *online* à avoir démontré l'influence d'indices phonétiques dans l'assignation perceptive d'un segment au mot qui précède ou qui suit est celle d'amorçage par répétition intermodale (Spinelli *et al.*, 2003). Contrairement à ce paradigme, le *word spotting* —bien qu'il présente sans doute d'autres désavantages— n'implique aucune

stimulation par l'écrit et donc garantit que les effets observés soient liés à l'influence directe des détails phonétiques sur l'activité lexicale, plutôt qu'à leur influence potentielle sur l'appariement phono-orthographique entre unités infra-lexicales (cf. Ziegler & Ferrand, 1998; Jacobs, Rey, Ziegler, & Grainger, 1998). Par ailleurs, l'emploi de cette tâche *online* offre une plus grande similitude avec l'étude de Dumay *et al.* (2002).

Comme nous l'avons déjà mentionné au chapitre 2, section 2.1.3, les études sur l'assignation lexicale d'un segment (Cohen, 1987; Nakatani & Dukes, 1977; O'Connor & Tooley, 1964; Quené, 1992, 1993; Spinelli *et al.*, 2003, pour ne citer que les études perceptives) ont toutes eu recours à des séquences ambiguës dont l'alternative était probablement évidente pour le locuteur même naïf (e.g., "cale égale" vs. "cas légal"). On ne peut donc exclure la possibilité que les différences acoustiques à la base des effets de segmentation rapportés soient la résultante d'un marquage articulatoire idiosyncratique et stratégique de la frontière entre les deux composants lexicaux. Et le fait que les stimuli aient été produits en contexte de phrase plutôt qu'isolément n'y change rien. Autrement dit, les phénomènes mis en évidence pourraient concerner uniquement les situations assez rares où le locuteur se trouve confronté à l'une des alternatives d'une paire minimale (e.g., "cas légal"), et ne pas s'appliquer aux frontières lexicales non-ambiguës (e.g., "cas léger"). Afin de contourner ce problème et d'évaluer plus à fond la généralité des indices des frontières de mots, les séquences à segmentation ambiguë utilisées dans cette thèse ont systématiquement été obtenues à partir d'énoncés non-ambigus du point de vue lexical. Ainsi, par exemple, les séquences d'une paire minimale telle que "tant#rou" vs. "tan#trou", dans lesquelles la frontière lexicale se situe soit à l'intérieur soit avant le groupe consonantique médian, étaient extraites d'énoncés non-ambigus tels que "tante roublarde" et "temps troublant", les portions "troublarde" et "roublant" ne correspondant à aucun mot en français. Cette méthodologie rend donc hautement improbable la possibilité que les différences acoustiques observées soient la conséquence directe de l'évocation d'une segmentation alternative au cours de la production des énoncés.

Par ailleurs, notons qu'ici, la structure grammaticale des énoncés d'origine, et partant les caractéristiques prosodiques induites par cette structure, ne variait pas d'une alternative à

l'autre (ni même d'une paire d'énoncés à l'autre), ce qui n'était pas le cas dans les études antérieures (Lehiste, 1960; O'Connor & Tooley, 1964; Nakatani & Dukes, 1977; Quené, 1992, 1993), où ce facteur contaminant est systématiquement présent. Selon les situations examinées dans la présente recherche, les énoncés utilisés étaient des types "nom + adjectif" ou "adjectif + nom".

3 Survol de la contribution expérimentale

Les expériences qui composent le **chapitre 4** auront pour but d'évaluer la présence éventuelle d'indices acoustiques/phonétiques sur la position des frontières de mots dans les groupes consonantiques OBstruante + LIquide (OBLI; cf. Dell, 1995) du français, et déterminer leur fonction perceptive dans la segmentation lexicale. L'**Expérience 1** étudiera les effets de l'alignement initial (e.g., "ic#roche" vs. "i#croche"; le soulignement indiquant les mots cibles) et final (e.g., "tant#rou" vs. "tan#trou") entre mots et frontières lexicales intentionnelle à l'aide du *word spotting*. Etant donné les résultats positifs de cette manipulation, les Expériences 2 et 3 chercheront à obtenir des données convergentes au moyen de tâches de segmentation *offline*. L'**Expérience 2** impliquera toujours la reconnaissance des mots, puisque la tâche des participants sera de catégoriser individuellement les séquences porteuses de l'Expérience 1 en fonction de la segmentation lexicale perçue: soit C#C soit #CC. En revanche, la tâche utilisée dans l'**Expérience 3** ne nécessitera pas la reconnaissance des mots, mais la syllabation explicite, les participants ayant à répéter ce qui constitue pour eux soit la première partie, soit la seconde partie des séquences porteuses (e.g., "croche" comme partie 2 pour "icroche"). Ce chapitre se clôturera par une **analyse corrélationnelle**, se servant notamment de la tâche de répétition partielle comme une sorte de "révélateur photographique" de la fonction perceptive des indices de frontières lexicales mis en évidence dans les deux autres tâches.

Le **chapitre 5** abordera la question de la nature générale (voire généralisable) vs. spécifique des variations phonétiques observées sur les groupes OBLI et des effets perceptifs qu'elles induisent, en comparant les OBLI à l'autre classe de groupes consonantiques potentiellement concernés par l'enchaînement: les groupes /s/ + OBstruante (SOB).

L'**Expérience 4** examinera l'effet d'alignement final en fonction de la nature du groupe consonantique à l'aide du *word spotting* (e.g., "ras#tu" vs. "ra#stu"). Cherchant à confirmer les résultats de détection lexicale, l'**Expérience 5** emploiera un paradigme d'appariement co-déterminé. Dans ce dernier, la tâche des participants sera de choisir parmi les deux alternatives auditives de chaque paire (e.g., "ras#tu" vs. "ra#stu") celle qui correspond le mieux à la segmentation cible, spécifiée au travers d'un énoncé visuel (e.g., "race tuméfiée" vs. "rat stupéfait"). L'**Expérience 6** étudiera le caractère systématique des variations de durée mises en évidence dans les expériences précédentes, sur base d'un large corpus produit par huit locuteurs naïfs. Enfin, l'**Expérience 6b** comparera l'enchaînement (e.g., "tant#rou") à une situation où la première consonne du groupe médian ne peut être resyllabifiée avec la consonne suivante (e.g., "tant#cou"). La comparaison entre ces deux conditions nous permettra de dire si la resyllabation dans le cas de l'enchaînement est juste incomplète, ou si elle n'a tout simplement pas lieu.

Les expériences rapportées aux **chapitre 6** seront les dernières de cette contribution. Elles auront pour objectif de déterminer les liens entre les variations acoustiques/phonétiques observées jusque là et le processus d'activation et de compétition lexicale. Plus particulièrement, l'**Expérience 7** examinera dans quelle mesure l'effet de non-alignement final en *word spotting* dépend de la présence d'un compétiteur (e.g., "trou") aligné à l'attaque de la syllabe suivante (e.g., "tan#trou" (vs. "tant#rou") vs. "tan#treum" (vs. "tant#reum"). Poursuivant dans cette voie, l'**Expérience 8** tentera d'élucider les relations entre effet d'alignement final, compétiteur enchâssant, et compétiteur du compétiteur, à l'aide des cibles lexicales les plus courtes (e.g., "tant#rou" vs. "tant#reum"). Comme les pièces d'un puzzle, les résultats de ces huit expériences permettront de statuer sur l'occurrence d'indices acoustiques/phonétiques codant la position des frontières des mots, et de mieux comprendre leur fonction perceptive dans la dynamique de la segmentation lexicale. Le **chapitre 7** conclura cette contribution en reprenant les principaux résultats et leurs implications théoriques.

I don't give a damn for a man that can only spell a word one way.

Mark Twain (1835-1910)

Chapitre 4 Indices acoustico-phonétiques des frontières de mots: expériences sur les groupes obli

D'après la conception de l'encodage phonologique proposée par Levelt et coll. (Levelt, 1989; Levelt et al., 1999; Levelt & Wheeldon, 1994), la parole ne devrait contenir en surface aucune trace des frontières entre les constituants lexicaux appartenant au même empan articulatoire (voir chapitre 3 pour plus de détails). Le but principal de ce chapitre est de mettre à l'épreuve cette prédiction. A cette fin nous avons comparé sur le front de la production, et sur celui de la perception de la parole, des paires de séquences auditives, dont l'une était un cas d'enchaînement qui impliquait un groupe consonantique (e.g., "tant#rou" extrait de tante roublarde), et l'autre, une séquence homophonémique dans laquelle le groupe médian était à l'attaque initiale du second mot (e.g., "tan#trou" extrait de temps troublant).

Les groupes consonantiques examinés, ici, appartenaient tous à la classe OBLI (cf. Dell, 1995). Il s'agit d'un sous-ensemble des suites obstruante-liquide, incluant uniquement les combinaisons d'une plosive ou des fricatives /f/ ou /v/ avec les liquides /ʀ/ ou /l/, et excluant /tl/ et /dl/, illégales en début de mot. Comme Goslin et Frauenfelder (2000) l'ont montré, la spécificité des groupes OBLI est qu'ils sont les seuls groupes consonantiques du français considérés de façon unanime comme indivisibles par les théories de la syllabation, que celles-ci soient basées sur le cycle de sonorité (Coursil, 1992; Peereman, 1999) ou sur les caractéristiques distributionnelles (Dell, 1995; Laporte, 1993). Pour cette raison, les groupes OBLI constituent donc un cas d'étude idéal pour évaluer l'hypothèse de resyllabation aux frontières de mots.

La première expérience de ce chapitre a évalué l'influence de l'alignement entre débuts (ou fins) de mots et frontières lexicales intentionnelles sur la segmentation *online*. La tâche utilisée était le *word spotting* (Cutler & Norris, 1988). L'effet d'alignement initial était examiné à l'aide de cibles de structure CVC (e.g., "roche") enchâssées en finale de porteurs VCCVC (e.g., "ic#roche" vs. "i#croche"); l'effet d'alignement final était examiné aussi à l'aide de cibles de structure CVC enchâssées cette fois à l'initiale de porteurs CVCCV (e.g.,

"tant#rou" vs. "tan#trou"). En outre, une analyse des durées segmentales visaient à mettre en évidence les différences acoustiques éventuelles entre les séquences #CC et C#C.

Expérience 1 – Effets d'alignement initial et final en word spotting

Méthode

Participants. Trente-deux étudiants (âge moyen: 21 ans, limites: 19-28) de l'Université libre de Bruxelles ont participé à l'expérience comme volontaires ou pour l'obtention de crédits dans le cadre d'un cours d'introduction à la psycholinguistique. Tous étaient de langue maternelle française et ont rapporté ne souffrir d'aucun trouble de l'audition ou du langage.

Matériel. L'effet d'alignement initial impliquait 32 mots de structure CVC enchâssés comme cibles en finale de bisyllabes VCCVC sans signification (e.g., "roche" dans "icroche") et dont la portion finale CCVC était aussi un mot (i.e., "croche"). Les cibles étaient toutes des noms ou adjectifs, d'une fréquence logarithmique moyenne de 2.97 (limites: 1.58-4.03) selon Brulex (Content, Mousty, & Radeau, 1990). Les groupes consonantiques médians appartenaient tous à la classe OBLI et consistaient en /bl, /gl/, /fl/, /pl/, /gʀ/, /dʀ/, /bʀ/, /tʀ/, /kʀ/, /fʀ/ et /pʀ/. Seize séquences avaient une obstruante voisée, suivie de /ʀ/ dans la moitié des cas et de /l/ dans l'autre; les séquences à obstruante sourde se caractérisaient par la même proportion de /ʀ/ et de /l/.

Deux versions auditives de chaque porteur ont été obtenues. Elles étaient extraites d'énoncés appariés de trois syllabes, comprenant deux mots et différant par la position de la frontière lexicale: …VC#CVC ou …V#CCVC. Afin d'éviter toute variation prosodique liée à la syntaxe, les énoncés étaient de même structure grammaticale (i.e., "adjectif + nom"). Dans la condition d'alignement, la cible était délibérément produite de sorte que son segment initial était aligné avec la frontière lexicale intentionnelle (e.g., "ic#roche" extrait de *magique roche*). En revanche, dans la condition de non-alignement, la cible émergeait, enchâssée dans le second mot, de sorte que son début ne correspondait pas avec la frontière lexicale intentionnelle (e.g., "i#croche" extrait de *demi-croche*). Afin d'éviter tout biais lexical sur la segmentation, la portion (i.e., V# vs. VC#) précédant chacun des deux points de segmentation

étudiés formait toujours un mot (e.g., "y" et "hic" dans "icroche"). Par ailleurs, la portion VCC n'était jamais un mot.

Trente-deux séquences VCCVC, construites selon les mêmes principes que les séquences porteuses si ce n'est que leur portion CVC ne correspondait à aucun mot, ont servi de remplisseurs (e.g., "ifraise" extrait de *actif raise* vs. *demi-fraise*). La distribution des groupes consonantiques au sein des remplisseurs était calquée sur celle caractérisant les porteurs. Ainsi, il n'y avait aucune corrélation entre l'identité de la suite de consonnes et l'occurrence d'une cible.

De façon similaire, l'effet d'alignement final impliquait 32 mots CVC (noms ou adjectifs; fréquence logarithmique moyenne: 3.61, limites: 2.59-5.01) enchâssés comme cibles à l'initiale de bisyllabes CVCCV sans signification (e.g., "tante" dans "tantrou"), avec la contrainte que la portion CV initiale était aussi un mot (e.g., "temps"). Ici, les groupes consonantiques OBLI utilisés étaient /pl/, /kl/, /bl/, /gl/, /tʀ/, /dʀ/ et /vʀ/.

A nouveau, deux versions auditives de chaque porteur ont été extraites de paires d'énoncés de trois ou quatre syllabes, comprenant deux mots et différant par la position de la frontière lexicale (e.g., "tant#rou" extrait de *tante roublarde* vs. "tan#trou" extrait de *temps troublant*); et 32 remplisseurs CVCCV ne contenant aucun mot CVC ont été calqués sur les séquences porteuses (e.g., "jufran" extrait de *juffe rembourrée* vs. *jus framboisé*). A l'exception du fait que l'adjectif ne précédait pas le nom mais le suivait, les séquences CVCCV étaient analogues aux séquences VCCVC.

L'ensemble du matériel —repris à l'Appendice 2— a été produit en chambre sourde par un locuteur masculin de langue maternelle française, à savoir l'auteur. Il avait pour consigne de lire silencieusement chaque énoncé avant de le prononcer à voix haute en enchaînant les deux termes, comme s'il le disait à quelqu'un. Les productions étaient capturées à l'aide d'un microphone Neumann U-87, et stockées sur DAT à une fréquence d'échantillonnage de 44.1 kHz et un taux de conversion D/A de 16 bits à l'aide d'un enregistreur Sony PCM-2300.

Après transfert des stimuli à une fréquence d'échantillonnage de 22.05 kHz sur le disque dur d'un Macintosh, porteurs et remplisseurs ont été extraits des énoncés grâce au logiciel Sound Designer II. Le découpage était tel qu'il n'incluait pas dans l'extrait la portion vocalique

contenant les indices sur l'identité de la consonne adjacente (e.g., la partie de /i/ affectée par /m/ dans *demi-croche*), cette dernière pouvant varier entre les énoncés d'une même paire. Lorsqu'elle était trop courte, la portion restante de cette voyelle était rallongée en clonant une de ses périodes, choisie dans la partie la plus stable, de sorte que la différence de durée de cette voyelle entre les deux versions alternatives ne soit pas supérieure à 10 ms. Les séquences auditives étaient ensuite normalisées afin d'éviter toute différence grossière d'amplitude.

Deux listes comprenant 64 porteurs et 64 remplisseurs en ordre pseudo-aléatoire fixe ont été composées. Toutes deux se scindaient en deux parties équilibrées, destinées respectivement à l'enchâssement initial et à l'enchâssement final. La position des cibles et des remplisseurs était identique d'une liste à l'autre, et plus de trois porteurs ou trois remplisseurs n'y apparaissaient jamais en succession directe. En revanche, la condition d'alignement dans laquelle se trouvait une cible particulière était contre-balancée à travers les listes, de sorte que chaque liste comportait 16 cibles par condition d'alignement et pour chaque type d'enchâssement. Les deux parties d'une liste se subdivisaient à leur tour en deux blocs équilibrés. Enfin, pour chaque type d'enchâssement, une liste de 30 essais d'un matériel similaire à celui des parties expérimentales, comprenant 20 porteurs dont 15 non-alignés, était destinée à servir de bloc d'entraînement.

Procédure. Les participants ont été testés individuellement dans une pièce isolée. La présentation des stimuli, le timing de l'expérience et le recueil des données étaient contrôlés au moyen du logiciel PsyScope 1.2.4 (Cohen, MacWhinney, Flatt, & Provost, 1993; voir aussi Appendice 7) et d'un Macintosh Power PC 7500/100. Les séquences de parole, stockées sur le disque dur de l'ordinateur, étaient présentées en monophonie binaurale à un volume d'écoute confortable, au moyen d'un amplificateur JVC JA-SG et d'écouteurs Beyerdynamic DT-770.

Les participants avaient pour consigne de presser un bouton aussi vite que possible (avec leur index préféré) dès qu'ils détectaient un mot enchâssé en position initiale (ou finale, selon la partie de l'expérience), et de répéter d'une voix intelligible le mot détecté. Dans la situation d'enchâssement final, ils ne pouvaient détecter que les mots débutant par une et une seule consonne (e.g., "roche" mais pas "croche"); dans celle d'enchâssement initial, ils ne pouvaient

détecter que les mots se terminant par une et une seule consonne (e.g., "tante" mais pas "temps")[17]. Les séquences de parole étaient précédées d'un signal d'avertissement de 100 ms présenté 1100 ms plus tôt. Pour la situation d'enchâssement final, la fin de la séquence auditive était séparée du début de l'essai subséquent par un intervalle de 3 s; pour celle d'enchâssement initial, cet intervalle prenait cours avec le début de la séquence et durait 3.5 s. Dans l'un et l'autre cas, il correspondait à la période d'enregistrement du temps de réaction. L'exactitude des productions orales des participants était contrôlée au vol par l'expérimentateur.

Comme les deux types d'enchâssement faisaient l'objet de parties distinctes et que les parties se scindaient à leur tour en deux blocs, l'ordre de passation des parties, ainsi que l'ordre des blocs dans chaque partie, était contre-balancé à travers les participants. Pour chaque type d'enchâssement, la familiarisation à la tâche impliquait deux présentations successives du bloc d'entraînement. Entre les blocs expérimentaux, une pause d'une minute était proposée. Une session complète durait environ 40 minutes. Seize participants ont été assignés à chaque liste.

Résultats et discussion

Dans la situation d'enchâssement initial, les temps de réaction ont été ajustés en soustrayant la durée du CVC afin d'être calculés à partir de la fin de la cible, comme dans la situation d'enchâssement final. Dans les deux cas, la mesure du temps de réaction était donc réalisée à partir de l'endroit du signal où, théoriquement, toute l'information pertinente pour la détection de la cible est disponible (cf. McQueen, 1996). Les analyses chronométriques étaient basées sur les temps de réaction accompagnés d'une réponse orale correcte. La distribution des temps de réaction pour chaque condition est rapportée à l'Appendice 3 (Figure 9). Dans l'ensemble, les taux d'erreurs moyens (i.e., omissions et réponses manuelles sans réponse orale ou accompagnées d'une réponse orale incorrecte) s'élevaient à 13.7% pour la détection en finale, et à 14.9% pour la détection à l'initiale. Les taux de fausses alarmes

[17] Bien que la portion CVCC ne pouvait de toute façon jamais être un mot, cette consigne, en spécifiant "se terminer par une et une seule consonne", était donc le miroir exact de celle utilisée pour l'enchâssement final.

étaient respectivement de 8.4% et 5.7%. Les temps de réaction moyens (par participant) et les taux d'erreurs moyens sont présentés au Tableau 1.

Tableau 1. Temps de réaction moyens (TRs) mesurés à partir de la fin de la cible, et taux d'erreurs moyens (Ers) pour l'Expérience 1. Les écarts-types figurent entre parenthèses.

Enchâssement final	Type de Porteur	TRs (ms)	Ers (%)
Alignée	ic#roche	920 (315)	13.5 (10.4)
Non-alignée	i#croche	1023 (286)	13.9 (14.3)
Différence		103 *pi*	0.4
Enchâssement initial			
Alignée	tant#rou	965 (349)	13.7 (11.2)
Non-alignée	tan#trou	1125 (305)	16.2 (14.5)
Différence		160 *pi*	2.5

Note. "*pi*" indique une différence statistiquement significative par participant et par item.

Comme on peut voir, les cibles enchâssées en finale de porteur étaient en moyenne détectées 103 ms plus tôt lorsque leur début était aligné avec les intentions lexicales du locuteur que lorsqu'il ne l'était pas. De la même façon, les cibles initiales étaient détectées 160 ms plus rapidement lorsque leur fin était alignée de la sorte que lorsqu'elle ne l'était pas. Des analyses de variance (ANOVAs) pour mesures répétées incluant l'alignement et la position d'enchâssement comme facteurs ont confirmé que l'effet de l'alignement entre mots et intentions lexicales était hautement significatif ($F_1(1,31) = 31.42$, $MSE = 17768.1$, $p < .001$; $F_2(1,62) = 27.33$, $MSE = 20877.7$, $p < .001$). En outre, comme en témoigne l'interaction entre alignement x position d'enchâssement marginalement significative par participant ($F_1(1,31) = 3.10$, $MSE = 8510.6$, $p = .09$; $F_2 < 1$), l'effet d'alignement montrait une tendance à être plus important pour les cibles initiales. Par ailleurs, la propension des cibles finales à être détectées plus rapidement (en moyenne 74 ms) que les cibles initiales était aussi marginalement significative par participant ($F_1(1,31) = 3.47$, $MSE = 50061.2$, $p = .072$; $F_2(1,62) = 1.91$, $MSE = 87535.5$, $p > .15$).

Hormis la légère tendance non-significative des cibles initiales à être détectées moins souvent dans la condition non-alignée que dans la condition alignée ($F_1(1,31) = 1.64$, $MSE = 62.8$, $p > .2$; $F_2(1,62) = 1.17$, $MSE = 88.1$, $p > .2$), les taux d'erreurs étaient sensiblement les mêmes à travers les conditions. Des ANOVAs similaires à celles effectuées sur les temps de réaction ne révélaient aucun effet ou interaction significative (tous les $Fs < 1$).[18]

La mise en évidence d'un coût perceptif sur les latences de détection à la fois pour les cibles finales et initiales suggère fortement que dans les groupes OBLI du français des variations liées à l'assignation lexicale de l'obstruante sont disponibles et exploitées *online* dans la segmentation. Par ailleurs, il est intéressant de remarquer que, contrairement aux résultats de Dumay *et al.* (2002), où seul le non-alignement de début entre mots et syllabes engendrait un coût perceptif, l'effet d'alignement final obtenu ici est d'une magnitude similaire (si pas plus importante) à celle de l'effet d'alignement initial.

Tableau 2. Durées segmentales moyennes (ms) en fonction de la structure phonologique du porteur et de la position de la frontière lexicale (#). Les écarts-types figurent entre parenthèses.

Conditions	C_1	V_1	OB	LI	V_2	C_4
ic#roche		76 (22)	110 (42) ***	97 (29) ***	149 (70)	253 (79) *
i#croche		68 (18)	94 (23)	62 (21)	146 (59)	227 (58)
tant#rou	111 (49)	134 (38) ***	81 (23) ***	63 (20) ***	62 (19)	
tan#trou	111 (39)	104 (30)	69 (19)	50 (16)	60 (18)	

Note. Une cellule représente 32 données. Les durées des segments affectés par la procédure d'extraction des stimuli sont rapportées en gris. Les différences significatives en fonction de la position de la frontière lexicale sont notées "*" si $p < .05$, "**" si $p < .01$, et "***" si $p < .001$.

Afin de déterminer la nature des différences acoustiques/phonétiques entre les deux versions auditives de chaque porteur, la durée de leur segments phonémiques a été estimée sur base à la fois de l'oscillogramme et de l'information parvenant à l'oreille. Cette analyse

[18] Par ailleurs indépendamment de l'alignement, il y avait dans le cas de l'enchâssement final une corrélation négative significative sur les latences ($r(32) = -.37$, $p < .04$), et marginalement significative sur les taux erreurs ($r(32) = -.34$, $p = .057$), entre la performance et la fréquence logarithmique de la cible, ce qui reproduit les résultats de Freedman (1992). En revanche, les corrélations correspondantes dans le cas de l'enchâssement initial étaient de direction opposée et non-significatives ($r(32) = .27$, $p > .13$ sur les TRs; $r(32) = .14$, $p > .4$ sur les Ers).

révélait des variations systématiques en fonction de la position de la frontière lexicale. En l'occurrence, les séquences C#C différaient des séquences #CC par des allongements substantiels de la première voyelle (des items CVCCV: 28.8%), de l'obstruante (17.2%), de la liquide (41.2%) et de la consonne finale (des items VCCVC: 11.5%). Les données sont présentées au Tableau 2.

Les statistiques étayant ces résultats provenaient d'ANOVAs menées séparément pour chaque segment. Dans le cas de l'obstruante (OB) et de la liquide (LI), celles-ci incluaient la structure phonologique des items et la position de la frontière lexicale comme facteurs. Pour les autres segments, dont V_1 et V_2 pour lesquels seules les données "non-rabotées" (i.e., les items C<u>V</u>CCV pour V_1 et VCC<u>V</u>C pour V_2) ont été prises en compte, seul le facteur "position de la frontière" était utilisé. L'effet de la position de la frontière était hautement significatif sur la durée de V_1 ($F(1,31) = 33.73$, $MSE = 430.5$, $p < .001$) et sur celle de l'obstruante ($F(1,62) = 24.70$, $MSE = 397.0$, $p < .001$), où il n'était pas modulé par la structure phonologique des items (tous les $Fs < 1$). Dans le cas de la liquide, l'effet de la position de la frontière était plus marqué pour les items VCCVC (56.5% vs. CVCCV: 26.0%), comme l'indiquait l'interaction position de la frontière x structure phonologique ($F(1,62) = 15.81$, $MSE = 225.9$, $p < .001$). Comme le démontraient des comparaisons planifiées, cet effet était significatif cependant pour les deux types de structure ($F(1,62) = 84.07$, $MSE = 225.9$, $p < .001$; $F(1,62) = 12.57$, $MSE = 225.9$, $p < .001$, respectivement). Enfin, l'effet de la position de la frontière était significatif sur la consonne finale des porteurs VCCVC ($F(1,31) = 4.83$, $MSE = 2344.9$, $p < .04$). En revanche, les segments C_1 et V_2 ne montraient pas d'effet (tous les $Fs < 1$). Ces mesures acoustiques suggèrent donc que des séquences ne différant que par l'assignation lexicale d'une consonne, soit initiale (e.g., i#<u>c</u>roche) soit finale enchaînée (i<u>c</u>#roche), ne sont pas homophones.

Comme on peut le voir au Tableau 2, ces variations de durée sur les segments ont pour corollaire que les cibles alignées étaient plus longues (de 53 ms) que les cibles non-alignées, comme l'atteste une ANOVA incluant position de la frontière et structure phonologique ($F(1,62) = 64.77$, $MSE = 1626.4$, $p < .001$). On pourrait donc penser que les résultats de *word spotting* soient dus, non pas à l'influence des corrélats phonétiques des frontières

syllabiques/lexicales sur l'activation des mots, mais au simple fait que les cibles alignées bénéficiaient d'un temps d'input plus long, et donc atteignaient potentiellement un niveau d'activation plus important. Plusieurs éléments indiquent cependant que l'interprétation des effets en termes d'alignement phonétique est l'interprétation qu'il faut privilégier. Premièrement, invoquer des différences de temps d'input ne permet pas de rendre compte de la tendance marginale de l'effet d'alignement à être plus important à l'initiale (e.g., tan#trou) qu'en finale de porteur (e.g., i#croche), puisque les différences de durée de cible montrent la tendance inverse (42 ms vs. 64 ms; $F(1,62) = 1.03$, $MSE = 1626$, $p > .3$). Deuxièmement, et sur le même front, nous avons calculé les corrélations entre l'effet d'alignement et la différence de durée de cible pour chaque type d'enchâssement (cf. Dumay *et al.*, 2002). Contrairement à l'explication ci-dessus qui prédit des corrélations négatives, les corrélations obtenues étaient de sens variable, faibles et non-significatives (VCCVC: $r(32) = -18$, $p > .3$ sur les TRs; $r(32) = .10$, $p > .5$ sur les Ers; CVCCV: $r(32) = -.05$, $p > .7$ sur les TRs; $r(32) = .07$, $p > .6$ sur les Ers). Enfin, le chapitre 6 présente deux résultats totalement incompatibles avec cette explication. L'un est que la différence de temps de réaction entre les conditions CVC#CV et CV#CCV dans la détection à l'initiale s'inverse lorsqu'on passe des cibles CVC aux cibles CV. Puisque les cibles CV sont, elles, plus courtes alignées que non-alignées, on voit mal comment des différences de temps d'input pourraient expliquer l'avantage systématique de la condition alignée (quel que soit le format des cibles). L'autre, nous nous garderons de le dévoiler ici. Etant donné ces éléments, on peut donc conclure que l'explication des effets en termes de différence entre les temps d'input des portions cibles n'est pas viable.

Aussi, de manière à approfondir les liens entre durées segmentales et performance de *word spotting*, nous avons examiné les corrélations entre les différences de durée sur la première voyelle, l'obstruante et la liquide, et l'effet d'alignement (i.e., C#C moins #CC) sur les latences de détection. Dans la tâche de détection à l'initiale (i.e., CVCCV), toutes les corrélations étaient non-significatives (V_1: $r(32) = .13$, $p > .4$; obstruante: $r(32) = -.25$, $p > .15$; liquide: $r(32) = .28$, $p > .1$). En revanche, pour la détection à la finale (i.e., VCCVC), il y avait une corrélation négative significative et substantielle entre la différence de durée sur la liquide et l'effet de l'alignement initial ($r(32) = -.40$, $p < .025$). Plus l'allongement de la liquide, de sa

variante non-initiale (#OBLI) à sa variante initiale (OB#LI), était grand, plus l'extraction du mot cible en condition alignée était rapide par rapport à la condition non-alignée (voir Figure 3). Les autres corrélations étaient, quant à elles, loin du seuil de signification (V_1: r(32) = -.25, $p > .15$; obstruante: r(32) = .25, $p > .2$). Bien qu'elle n'en constitue pas une preuve irréfutable, l'obtention de cette corrélation sur la liquide est fortement suggestive de l'ancrage de la performance de segmentation dans les variations des durées segmentales. Par ailleurs, le fait que cette corrélation n'émerge que dans la situation d'enchâssement final indique que parmi ces variations, celle qui touche le segment initial de la représentation lexicale est la plus déterminante.

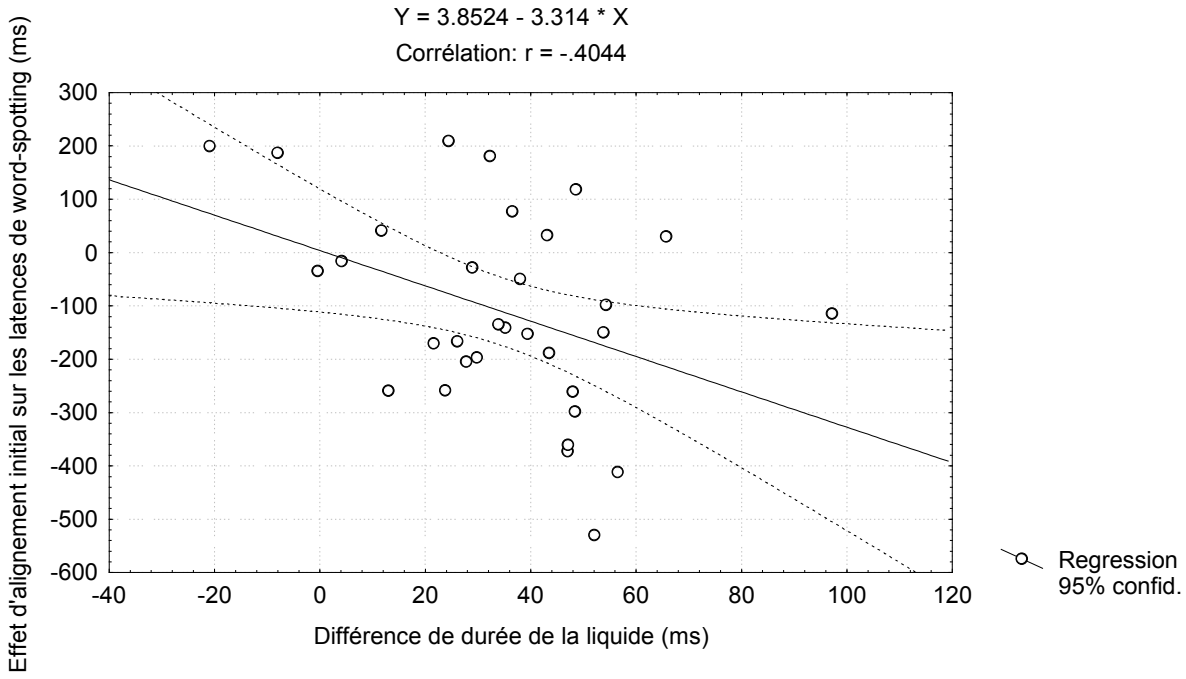

Figure 3. Corrélation (et droite de régression) entre l'effet d'alignement initial (e.g., ic#roche vs. i#croche) sur les latences de *word spotting* et la différence de durée de la liquide entre les conditions C#C et #CC.

Les résultats de cette première expérience corroborent ceux de deux études-clés. A l'aide d'une tâche de catégorisation, Quené (1992; voir aussi Quené, 1993) a démontré que les auditeurs néerlandophones exploitent la durée de la consonne pivotale et le temps de montée

en énergie de la voyelle subséquente pour distinguer les consonnes intervocaliques enchaînées de leurs variantes initiales (e.g., "diep in" vs. "die pin"). Au moyen de l'amorçage par répétition intermodale, Spinelli *et al.* (2003) ont montré que si les consonnes de liaison du français activent une cohorte initiale, l'activation lexicale produite n'est pas à la hauteur de celle générée par leurs variantes de début de mot. Nos résultats confirment ceux de Quené à l'aide d'une tâche de segmentation *online* et de paires d'énoncés de même structure grammaticale, et les étendent aux groupes consonantiques OBLI du français. Ils corroborent ceux de Spinelli *et al.* à l'aide d'un autre cas de resyllabation potentielle —l'enchaînement— et d'un paradigme de segmentation *online* différent.

SOSH, l'hypothèse de segmentation aux attaques syllabiques (Content *et al.*, 2001; Dumay *et al.*, 2002; Frauenfelder & Content, 1999), rend directement compte de l'effet d'alignement initial (e.g., ic#roche vs. i#croche). Dans la condition non-alignée, le point de segmentation lexicale le plus probable se situe vraisemblablement sur le segment précédent. Or comme l'accès qui s'y opère ne conduit pas à reconnaître une cible valide (e.g., "roche"), l'auditeur doit revoir ses plans, et cela prend du temps. En revanche, SOSH ne prédit directement aucun effet d'alignement final (e.g., "tant#rou" vs. "tan#trou"), puisque selon cette hypothèse seul l'alignement entre le début du mot et le début d'une syllabe est important pour l'activation lexicale. Toutefois, étant donné les caractéristiques lexicales des séquences porteuses, il est fort probable qu'en position initiale, la cible entre systématiquement en compétition avec le mot CCV (e.g., "trou") pour l'assignation lexicale de la consonne pivotale (e.g., "t"). Puisque dans la condition non-alignée, le compétiteur débute sur une attaque de syllabe et est donc particulièrement activé, il s'approprierait plus facilement la consonne pivotale et, du même coup, il empêcherait davantage l'auditeur de percevoir la cible (e.g., "tant"). Sous cet angle, et en compatibilité avec les résultats de Dumay *et al.* (2002), la situation d'enchâssement initial ne montrerait pas un effet de non-alignement final au sens propre du terme, mais un effet rétroactif lié à l'alignement initial du compétiteur chevauchant. Le but de l'Expérience 7 sera de tester cette hypothèse.

Enfin pour clôturer cette discussion et de manière complémentaire avec les idées développées au paragraphe précédent, nous avons évalué par le biais de corrélations le lien

entre l'effet d'alignement initial (e.g., "ic#roche" vs. "i#croche") et l'influence inhibitrice du compétiteur enchâssant (e.g., "croche"). Les latences de détection révélaient une corrélation négative significative entre l'effet de l'alignement (C#C moins #CC) et la fréquence (logarithmique) du compétiteur CCVC ($r(32) = -.44$, $p < .02$)[19], en l'absence de corrélation (négative) entre la fréquence du compétiteur et la fréquence de la cible ($r(32) = .02$, $p > .9$) ou entre la fréquence de la cible et l'effet d'alignement ($r(32) = .11$, $p > .5$). Le non-alignement de la cible (e.g., "roche") apparaît donc d'autant plus difficile à surmonter que le compétiteur (e.g., "croche") évoqué en contre-partie est fréquent. L'Expérience 2 a cherché à corroborer les résultats de *word spotting* à l'aide d'une tâche de segmentation lexicale *offline*.

Expérience 2 – Perception des frontières de mots en tâche de catégorisation

La tâche était celle de catégorisation, tâche classiquement utilisée par O'Connor et Tooley (1964), Christie (1974) et plus récemment Quené (1992, 1993). Les participants avaient à classer les séquences porteuses de l'Expérience 1 comme appartenant à la condition C#C ou à la condition #CC. En ne présentant qu'une seule des deux alternatives auditives (e.g., "i#croche"), on évitait par ailleurs que les participants ne puissent les comparer l'une à l'autre, et que les réponses ne soient co-déterminées (voir chapitre 2, section 2.1.1.2). Sous l'hypothèse que les porteurs d'une même paire ne sont pas strictement homophones, on s'attendait à une performance de catégorisation supérieure au niveau du hasard (50%), avec peut-être un désavantage pour les stimuli VCCVC, les différences de durée sur leur première voyelle ayant été gommées au cours de l'extraction de ces stimuli à partir des énoncés.

Méthode

Participants. Vingt-quatre étudiants (âge moyen: 19 ans, limites: 16-20) de l'Université libre de Bruxelles ont participé à l'expérience comme volontaires ou en échange de crédits dans le cadre d'un cours d'introduction à la psychologie expérimentale. Tous étaient de langue

[19] Les corrélations correspondantes sur les erreurs étaient non-significatives (effet de l'alignement et fréquence (logarithmique) du compétiteur: $r(32) = .18$, $p > .3$; effet de l'alignement et fréquence de la cible: $r(32) = -.22$, $p > .2$).

maternelle française et ont rapporté ne souffrir d'aucun trouble de l'audition ou du langage. Aucun d'eux n'avait participé à l'Expérience 1.

Matériel. Le matériel auditif était composé des 64 paires de séquences VCCVC ou CVCCV utilisées comme porteurs dans l'Expérience 1 (e.g., "ic#roche" vs. "i#croche"; "tant#rou" vs. "tan#trou"). Ces séquences étaient réparties en deux listes expérimentales identiques à celles élaborées pour l'Expérience 1, si ce n'est que les remplisseurs en étaient exclus. En miroir des listes auditives (où à chaque essai une seule version d'un même porteur était présentée), deux listes visuelles ont été composées. Dans chaque liste visuelle, les deux énoncés d'origine (e.g., "magique roche" et "demi-croche") apparaissaient côte à côte. La moitié des énoncés de type C#C était présentée à gauche, et l'autre à droite, tandis que la position gauche vs. droite des énoncés d'une même paire était contre-balancée à travers les listes.

Procédure. Les participants ont été testés individuellement dans une pièce isolée. Les séquences auditives étaient présentées à un volume d'écoute confortable, au moyen d'un lecteur DAT Sony DTC-60ES et d'écouteurs Beyerdynamic DT-770. Les paires d'énoncés étaient présentées par écrit en une colonne centrée sur les feuillets d'un protocole, en omettant la partie non-pertinente (e.g., "…ique roche" vs. "…i croche"). A chaque essai, les participants avaient pour tâche de choisir parmi les deux segmentations représentées visuellement celle qui correspondait le mieux à la séquence auditive. Chaque essai démarrait par un signal de 50 ms. A ce signal, les participants devaient évaluer les différences articulatoires mises en jeu par l'enchaînement des deux termes dans chacun des deux énoncés visuels. Au bout de 8 s, un second signal prévenait les participants de l'imminence de la stimulation auditive. Celle-ci débutait 800 ms plus tard et consistait en quatre présentations de la séquence porteuse, séparées les unes des autres par un intervalle de 2500 ms. Au terme de la quatrième présentation, les participants disposaient encore de 6 s pour répondre impérativement. Et revenir à un essai antérieur n'était pas permis.

Douze participants ont été assignés à chaque liste auditive (i.e., six pour chaque liste visuelle). Comme chaque moitié de liste était dédiée à une structure phonologique fixe, i.e., VCCVC ou CVCCV, elle était précédée d'un entraînement spécifique de huit essais similaires

aux essais expérimentaux. Contrairement à l'expérience de *word spotting*, tous les participants démarraient par les items VCCVC. Comme précédemment cependant, l'ordre de présentation des deux blocs constituant chaque moitié était contre-balancé à travers les participants. L'expérience durait environ 40 minutes.

Résultats et discussion

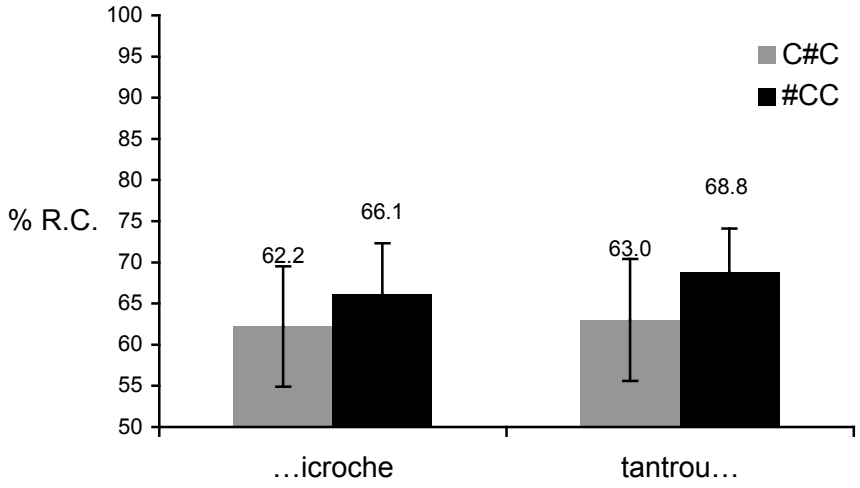

Figure 4. Taux moyen de réponses correctes (et intervalle de confiance) en fonction de la structure phonologique des stimuli et de la position de la frontière lexicale pour l'Expérience 2.

Les taux de réponses correctes en fonction de la structure phonologique des stimuli et de la position de la frontière lexicale sont présentés à la Figure 4. Comme on peut le voir, la catégorisation des séquences s'avérait particulièrement difficile, avec un taux global de seulement 65%. Néanmoins, comme l'indiquaient des tests-t unilatéraux, la performance des sujets était significativement supérieure au niveau du hasard (de 50%) par participant et item dans les quatre conditions (ic#roche: 62.2% (17.2), $t_1(23) = 3.48$; $p < .002$; $t_2(31) = 3.63$; $p < .001$; i#croche: 66.1% (14.7), $t_1(23) = 5.37$; $p < .001$; $t_2(31) = 4.41$; $p < .001$; tant#rou: 63.0% (17.6), $t_1(23) = 3.63$; $p < .001$; $t_2(31) = 3.84$; $p < .001$; tan#trou: 68.8% (12.5), $t_1(23) = 7.35$; $p < .001$; $t_2(31) = 8.24$; $p < .001$). Les influences de la position de la frontière lexicale et de la structure phonologique des stimuli ont été examinées au moyen d'ANOVAs pour mesures répétées. En dépit d'une légère tendance des items C#C à être (1.15 fois) plus souvent catégorisés erronément que leurs analogues #CC, l'effet de la position de la frontière lexicale

n'était pas significatif ($F_1(1,21) = 1.22$, $MSE = 456.0$, $p > .25$; $F_2(1,62) = 1.51$, $MSE = 491.1$, $p > .25$)[20]. Par ailleurs, bien que la différence entre les items VCCVC et les items CVCCV allait dans le sens prédit (64.2% pour les items VCCVC vs. 65.9% pour les items CVCCV), l'effet de la structure phonologique était loin du seuil de signification (les deux $Fs < 1$). Enfin, il n'y avait pas non plus d'interaction significative entre la structure phonologique et la position de la frontière lexicale (les deux $Fs < 1$).

Au regard de ces résultats, il est clair que les auditeurs ont beaucoup de difficulté à juger des intentions lexicales sous-jacentes aux séquences porteuses de l'Expérience 1. Néanmoins, étant donné la nature significative des tests-t par item, on peut en déduire que dans la majorité des paires de porteurs, des éléments liés à la position de la frontière lexicale sont disponibles et permettent de faire la distinction entre les deux stimuli l'un de l'autre. Par conséquent, la faible performance de catégorisation ne peut être interprétée en termes du fait que les indices des frontières lexicales OBLI seraient souvent de peu de poids voire absents, mais plutôt en termes d'une incapacité manifeste des auditeurs à les exploiter correctement et systématiquement dans cette situation "métaphonétique". Quoi qu'il en soit, l'obtention d'une performance supérieure au niveau du hasard, si petite soit-elle, corrobore sur les mêmes stimuli les résultats de l'Expérience 1 obtenus à l'aide du *word spotting*.

Comme précédemment, afin d'explorer les liens entre la performance et les variations des durées segmentales, nous avons calculé les coefficients de corrélation entre les différences de durée sur la première voyelle, l'obstruante et la liquide (i.e., C#C moins #CC), et la performance de catégorisation moyenne par paire de porteurs. Comme pour le *word spotting*, le seul segment des items VCCVC à montrer une corrélation (ici marginalement) significative était la liquide ($r(32) = .34$, $p = .057$). Plus /l/ ou /ʀ/ (en souligné) dans la condition VC#CVC était long par rapport à la condition V#CCVC, plus les intentions lexicales du locuteur avaient de chance d'être correctement perçues. Ici, les items CVCCV attestaient eux-aussi d'un lien entre la performance et les durées segmentales, puisqu'ils montraient une corrélation significative sur la première voyelle ($r(32) = .42$, $p < .02$), de même sens que celle mise en

[20] Cette légère asymétrie est vraisemblablement imputable à la structure phonotactique du français, puisque les syllabes ouvertes y prédominent.

évidence sur la liquide des items VCCVC. En dépit de l'absence d'un effet de la structure phonologique dans les ANOVAs, il semble donc que les variations de durée sur la première voyelle, gommées dans les items VCCVC mais préservées dans les CVCCV, soient elles-aussi pertinentes pour la segmentation *offline*. Les autres corrélations étaient non-significatives (V_1: r(32) = .23, $p > .2$; obstruante: r(32) = -.02, $p > .9$, pour les items VCCVC; obstruante: r(32) = -.26, $p > .15$; liquide: r(32) = -.05, $p > .8$, pour les items CVCCV).

Sur base des résultats de ces deux premières expériences, il est donc clair que les auditeurs exploitent des éléments d'ordre acoustico-phonétique présents dans les groupes OBLI, qui leur permettent de distinguer les cas d'enchaînement de l'obstruante (C#C), de ceux où cette dernière est à l'initiale du mot suivant (#CC). Cependant, étant donné la nature des tâches utilisées (i.e., *word spotting* et catégorisation), il est difficile de dire si l'exploitation de ces indices, notamment à des fins de modulation de l'activité lexicale, est indépendante de l'analyse lexicale intentionnelle des énoncés que nécessitent ces deux tâches. L'Expérience 3 a tenté de répondre à cette question en utilisant une tâche de segmentation orthogonale au traitement lexical.

Expérience 3 – Incidence des frontières de mots sur la syllabation explicite: données obtenues en tâche de répétition partielle

Le but de l'Expérience 3 était de tester l'influence des indices liés à l'assignation lexicale de l'obstruante sur les choix de syllabation explicite. La tâche, dite de "répétition partielle", introduite par Treiman, Gross et Cwikiel-Glavin (1992, Expérience 2), consistait à répéter soit la première partie soit la seconde partie de séquences auditives, en l'occurrence ici, les porteurs de l'Expérience 1. Aussi, afin d'augmenter nos chances de capturer l'influence de ces indices de bas niveau, et contrairement à sa variante classique, cette tâche était effectuée sous pression temporelle (cf. Goslin & Frauenfelder, 2000).

Comme l'ont montré Content et coll. (Content *et al.*, 2001; Content, Dumay, & Frauenfelder, 1999; Goslin, Content, & Frauenfelder, 1999; Goslin & Frauenfelder, 2000), dans le cas des singletons consonantiques médians et des groupes OBLI du français, les décisions d'ouverture (ou d'attaque) syllabique (impliquées dans la répétition de la seconde

partie) sont beaucoup plus cohérentes et moins labiles que celles de clôture/coda syllabique (impliquées dans la répétition de la première partie), très perméables à l'influence de variables phonologiques (e.g., sonorité de la consonne pivotale; aperture des voyelles adjacentes), lexicales (e.g., présence de mots enchâssés) ou encore orthographiques (e.g., redoublement de lettres). La présente expérience étant centrée sur les groupes OBLI, elle constituait donc une opportunité supplémentaire de reproduire cette dissociation.

Méthode

Participants. Trente-deux étudiants (âge moyen: 18 ans, limites: 18-21) de l'Université libre de Bruxelles ont participé à l'expérience, en échange de crédits dans le cadre d'un cours d'introduction à la psychologie expérimentale. Tous étaient de langue maternelle française et ont rapporté ne souffrir d'aucun trouble de l'audition ou du langage. Aucun d'eux n'avait participé aux expériences précédentes.

Matériel. Le matériel était composé des 64 paires de porteurs/stimuli utilisés dans les deux expériences précédentes: 32 paires d'items VCCVC (e.g., "ic#roche" vs. "i#croche") et 32 paires d'items CVCCV (e.g., "tant#rou" vs. "tan#trou"). Les stimuli étaient répartis en deux listes expérimentales contenant chacune une alternative de chaque paire, et la position de la frontière lexicale était contre-balancée, de sorte que chaque liste comportait 16 items C#C et 16 items #CC par structure phonologique. Par ailleurs, chaque liste se scindait en deux blocs en fonction de la structure phonologique des stimuli.

Procédure. Les participants ont été testés individuellement dans une pièce isolée. La présentation des stimuli, le timing de l'expérience et le recueil des données étaient contrôlés au moyen du logiciel PsyScope 1.2.4 (Cohen *et al.*, 1993) et d'un Macintosh-G3 Power PC. Les séquences de parole, stockées sur le disque dur de l'ordinateur, étaient présentées en monophonie binaurale à un volume d'écoute confortable, au moyen d'un amplificateur Kenwood KAF-3010R et d'écouteurs Beyerdynamic DT-770. Chaque séquence était précédée d'un signal de 50 ms qui débutait 550 ms plus tôt et prévenait les participants de l'imminence de la présentation. Ceux-ci avaient pour consigne de répéter dès que possible, et sans que la langue leur fourche, ce qui constituait pour eux soit la première, soit la seconde partie de la

séquence de parole. Pour être valide, une réponse de première partie devait nécessairement débuter par le segment initial de la séquence et ne pouvait contenir qu'une voyelle (e.g., "i", "ic", et "icr" pour "icroche"; "tan", "tant" et "tantr" pour "tantrou"); à l'inverse, outre le fait qu'elle ne pouvait contenir qu'une seule voyelle, une réponse de seconde partie devait nécessairement finir par le segment final de la séquence (e.g., "croche", "roche"et "oche" pour "icroche"; "trou", "rou" et "ou" pour "tantrou"). Tout ceci était expliqué à l'aide d'exemples et sans faire appel à la notion de syllabe à aucun moment. Les réponses étaient encodées au vol par l'expérimentateur, tandis que les latences étaient mesurées à l'aide d'un microphone situé à environ 6 cm des lèvres du participant, et relié à l'ordinateur via une clé vocale. Dès que celle-ci détectait la réponse, un temps de réaction calculé à partir du début de la séquence était enregistré, et un signal apparaissait à l'écran de l'expérimentateur. Cette stimulation visuelle lui permettait de vérifier que le déclenchement de la clé vocale coïncidait avec le début de l'articulation. L'intervalle entre les débuts de deux essais successifs dépendait directement du temps de réaction, de sorte que le rythme de l'expérience variait en fonction de la vitesse du sujet. Aussi, afin d'éviter toute collision entre la réponse et l'essai subséquent, une période fixe de 2 s séparait le début de l'articulation du signal d'avertissement subséquent. En cas de non-réponse endéans les 3 s qui suivaient le début de la séquence, l'essai était considéré comme manqué et le programme passait automatiquement à la période fixe, précédant l'essai subséquent.

Seize participants ont été assignés à chaque liste, huit d'entre eux débutant par le bloc d'items VCCVC, les autres débutant par le bloc d'items CVCCV. La liste était présentée deux fois, de sorte que les participants puissent effectuer les deux tâches sur chaque item. La passation des tâches suivait un ordre de type ABBA, la première tâche à réaliser (i.e., répétition de la première vs. de la seconde partie) étant déterminée par la structure phonologique des stimuli. Les participants d'abord confrontés aux items VCCVC commençaient par répéter la seconde partie de ces items, puis enchaînaient par la répétition de la première partie des items CVCCV. En miroir, les participants d'abord soumis aux items CVCCV commençaient par répéter la première partie de ces items, puis enchaînaient par la répétition de la seconde partie des items VCCVC. Chaque tâche faisait l'objet d'un

entraînement spécifique de 12 essais pour la première passation, et de six pour la seconde. Les participants bénéficiaient de deux essais de démarrage à chaque bloc. L'expérience durait un peu plus de 20 minutes.

Résultats et discussion

Tableau 3. Répartition des réponses en fonction de l'assignation syllabique des consonnes du groupe médian.

		.CC	C.C	CC.	Autre	Absente
Exemples	Partie 1	tan	tant	tantr	ant, …etc.	
	Partie 2	trou	rou	ou	antrou, …etc.	
	Proportion	.60	.39	.002	.003	.005

De manière à obtenir un codage des données indépendant de la partie à répéter, les réponses ont été catégorisées en fonction de l'assignation syllabique des consonnes du groupe médian. Ainsi, et pour autant qu'elles incluent, le cas échéant, le phonème initial (ou final) attendu sur base de la partie à répéter, les réponses …V]CCV[21]… comme première partie et …V[CCV… comme seconde partie ont été codées ".CC", les réponses …VC]CV… comme première partie et …VC[CV… comme seconde partie ont été codées "C.C", et enfin, les réponses …VCC]V… comme première partie et …VCC[V… comme seconde partie ont été codées "CC.".

Comme on peut le voir au Tableau 3, les proportions de réponses .CC et C.C étaient indéniablement complémentaires, puisque les deux catégories expliquaient à elles seules 99% des données. Par conséquent, les autres catégories de réponses, insignifiantes, ont été exclues des analyses, et la proportion de réponses (canoniques) .CC, par rapport au total des réponses .CC et C.C, a été utilisée comme variable dépendante.

[21] Le symbole "]" indique la fin de la réponse fournie comme Partie 1; le symbole "[" indique le début de la réponse fournie comme Partie 2.

Tableau 4. Proportion moyenne de réponses .CC (et écarts-types) en fonction de la partie à répéter, de la structure phonologique des stimuli et de la position de la frontière lexicale.

	Partie 1		Partie 2	
	VCCVC	CVCCV	VCCVC	CVCCV
C#C	.32 (.35)	.44 (.40)	.76 (.27)	.70 (.28)
#CC	.39 (.37)	.53 (.37)	.87 (.22)	.80 (.24)
M	.35 (.36)	.49 (.38)	.81 (.24)	.75 (.25)

Le Tableau 4 présente la proportion moyenne de réponses .CC par condition. Deux ANOVAs ont été effectuées, l'une sur les proportions calculées par participant, et l'autre sur celles calculées par item. Ces analyses incluaient trois variables indépendantes: la partie à répéter (première vs. seconde), la structure phonologique des stimuli (VCCVC vs. CVCCV) et la position de la frontière lexicale (C#C vs. #CC). A l'instar des études précédentes, la proportion de réponses canoniques (.CC) était bien plus importante en répétition de seconde partie, où elle était de .78, qu'en répétition de première partie, où elle tombait à .42 ($F_1(1,31) = 64.33$, $MSE = 1303.0$, $p < .001$; $F_2(1,62) = 453.60$, $MSE = 183.8$, $p < .001$). Autrement dit, les participants incluaient l'obstruante en attaque de la seconde partie (V[CC) deux fois plus souvent qu'ils ne l'excluaient de la rime de la première partie (V]CC). Plus important ici, l'effet de la position de la frontière lexicale était hautement significatif ($F_1(1,31) = 52.60$, $MSE = 93.3$, $p < .001$; $F_2(1,62) = 27.99$, $MSE = 179.0$, $p < .001$): les réponses .CC étaient moins fréquentes lorsque la frontière lexicale se situait entre les deux consonnes du groupe médian (.56) que lorsqu'elle les précédait (.65). En outre, contrairement aux variables manipulées dans les études précédentes, l'influence de la position de la frontière lexicale était clairement indépendante de la partie à répéter, comme en témoigne l'absence d'interaction partie x position (les deux $Fs < 1$). En moyenne, pour les deux de types de structure phonologique, la réduction observée dans la proportion de réponses .CC était de .8, pour la Partie 1, et de .10, pour la Partie 2 (cf. Tableau 4). Enfin, la structure phonologique des stimuli n'avait globalement pas d'influence significative ($F_1 < 1$; $F_2(1,62) = 1.79$, $MSE = 404.4$, $p > .15$), mais elle interagissait avec la variable "partie" ($F_1(1,31) = 8.11$, $MSE = 756.2$,

$p < .01$; $F_2(1,62) = 35.12$, $MSE = 183.8$, $p < .001$). Comme l'attestent des comparaisons planifiées, bien que l'effet de la partie à répéter était plus faible pour les items CVCCV (.49 vs. .75) que pour les items VCCVC (.35 vs. .81), il était significatif pour les deux types de structure phonologique (CVCCV: $F_1(1,31) = 25.80$, $MSE = 864.3$, $p < .001$; $F_2(1,62) = 118.1$, $MSE = 183.8$, $p < .001$; VCCVC: $F_1(1,31) = 56.62$, $MSE = 1194.9$, $p < .001$; $F_2(1,62) = 370.58$, $MSE = 183.8$, $p < .001$). Par ailleurs, la proportion de réponses .CC pour la Partie 1 tendait à être plus faible pour les stimuli VCCVC que pour ceux de structure CVCCV (.35 vs. .49; $F_1(1,31) = 3.54$, $MSE = 1544.7$, $p < .07$; $F_2(1,62) = 18.71$, $MSE = 307.3$, $p < .001$), les participants étant visiblement très réticents à donner une réponse qui ne comporte qu'un seul segment (i.e., V]CCVC vs. CV]CCV). En revanche pour la Partie 2, c'est la tendance opposée, mais non-significative par participant, qui s'observait (VCCVC: .82 vs. CVCCV: .75; $F_1(1,31) = 2.33$, $MSE = 582.4$, $p < .15$; $F_2(1,62) = 5.09$, $MSE = 281.0$, $p < .03$); ici, le fait que les réponses .CC aux items VCCVC laissaient inéluctablement de côté le segment vocalique initial (qu'ils étaient réticents à produire en tâche de Partie 1) n'avait manifestement aucune influence sur leurs choix de syllabation. Les analyses ne révélaient aucune autre interaction significative (tous les *Fs* < ou proches de 1). Etant donné la disproportion massive entre les réponses .CC et C.C, particulièrement dans la tâche de seconde partie, les latences de répétition n'ont pas été traitées.

En somme, les données de cette expérience de répétition partielle reproduisent incontestablement l'asymétrie mise en évidence par Content *et al.* (2001) entre les décisions d'ouverture de syllabe, hautement cohérentes, et celles de clôture syllabique, nettement plus variables. En l'occurrence, les participants montraient une tendance quasi-systématique à assigner l'obstruante et la liquide en attaque de seconde syllabe (e.g., "i[croche"), lorsqu'ils avaient à ouvrir la seconde partie de l'énoncé. En revanche, lorsqu'il s'agissait de déterminer la fin de la première partie, ils étaient franchement plus indécis et incohérents quant au sort syllabique de l'obstruante, produisant aussi bien, voire même plus souvent, "ic]roche" que "i]croche".

Plus pertinent pour la présente recherche, les décisions d'ouverture et de clôture syllabique étaient très clairement attirées, et ce avec la même force, vers la position de la frontière

lexicale dans les énoncés d'origine (e.g., "tante roublarde" vs. "temps troublant"). Que ce soit pour la Partie 1 ou la Partie 2, il y avait moins de réponses .CC pour la condition C#C (enchaînement; e.g., "tant#rou") que pour la condition #CC (e.g., "tan#trou"). Ce résultat confirme indéniablement la présence d'indices liés à l'assignation lexicale de l'obstruante dans les groupes OBLI, et corrobore du même coup les effets obtenus en *word spotting* et en catégorisation. L'analyse lexicale intentionnelle requise par ces deux tâches ne semble donc pas fondamentale à l'exploitation des indices en question.

Comme pour les Expériences 1 et 2, nous avons calculé les corrélations entre l'effet de la position de la frontière lexicale sur la proportion de réponses .CC et les différences sur les durées segmentales, selon la partie à répéter et la structure phonologique des items. Au contraire des données de *word spotting* et de catégorisation, aucune des corrélations examinées ici, que ce soit pour la première voyelle, l'obstruante, ou la liquide, n'était significative (tous les $r_s(32)$ entre .-29 et .17, tous les *ps* > .1). Par conséquent, il est difficile d'affirmer que l'effet de l'assignation lexicale de l'obstruante sur les choix de syllabation explicite est lui aussi imputable aux co-variations résultantes sur les durées segmentales.

Enfin, concernant l'asymétrie entre décisions d'ouverture et de clôture syllabique (cf. Content *et al.*, 2001), Goslin et Frauenfelder (2000) indiquent que la présence de l'accent sur la seconde syllabe dans les expériences en français pourrait expliquer l'assignation presque systématique des consonnes intervocaliques (et des groupes OBLI) en attaque de réponse de seconde partie (e.g., "vé[lo", "ci[tron"). Comme en anglais, où le *stress*, souvent initial, attire à lui les consonnes subséquentes *a priori* hétérosyllabiques, et crée des situations d'ambisyllabicité (e.g., "le[m]on" où le /m/ semble appartenir au deux syllabes à la fois; Kahn, 1976; Sapir, 1933; Selkirk, 1982; Treiman & Danis, 1988), en français, la valeur accentuelle de la partie au cœur de la fenêtre attentionnelle déterminerait sa capacité à attirer vers elle les consonnes adjacentes. Appliquée à nos stimuli, cette interprétation ne prédit donc pas de plus grande systématicité dans l'assignation de l'obstruante en attaque de syllabe en tâche de Partie 2 qu'en tâche de Partie 1, pour les items CVCCV (e.g., "tantrou"). Ces séquences étant extraites d'énoncés plus longs, composés d'un nom d'une syllabe et d'un adjectif de deux (e.g., "tante roublarde"), même relative, la valeur accentuelle de la seconde syllabe (e.g., "…ou")

est beaucoup plus faible que celle de la première (tan...). La mise en évidence d'une asymétrie significative entre première et seconde partie pour les items CVCCV, bien que légèrement plus faible pour ces items que pour ceux de structure VCCVC, va donc à l'encontre de cette interprétation. En somme, la cohérence plus grande pour les débuts que pour les fins de syllabes n'est pas réductible à une histoire d'accent.

Analyse corrélationnelle entre word spotting, catégorisation et répétition partielle

Les Expériences 1-3 montrent que les auditeurs francophones exploitent la présence de variations phonétiques liées à l'assignation lexicale de l'obstruante des groupes consonantiques OBLI dans des tâches de segmentation lexicale, mais aussi dans leurs choix de syllabation explicite. Une analyse acoustique des stimuli révèle par ailleurs que ces variations se manifestent notamment par des différences sur les durées segmentales aux abords de la frontière. Aussi, des corrélations entre la performance et les phénomènes de durée suggèrent fortement l'implication de la voyelle précédente (en tâche de catégorisation) et de la liquide subséquente (en tâche de *word spotting* et de catégorisation) dans la localisation de la frontière lexicale. Le but premier de cette section est de déterminer la fonction perceptive de ces phénomènes phonétiques. Plus précisément, elle tente de répondre à la question de savoir si les variations mises en évidence sont perçues comme des indices de frontière entre deux mots/syllabes ou, plus restrictivement, uniquement comme des indices de début de mot/syllabe, ou de fin de mot/syllabe. Notons que l'hypothèse selon laquelle ces variations sont perçues exclusivement comme des indices de début s'inscrit directement dans la même ligne que l'interprétation —en termes d'effet rétroactif inhibiteur du compétiteur chevauchant— donnée par SOSH (Content *et al.*, 2001; Dumay *et al.*, 2002) à l'effet de non-alignement final (en *word spotting*; e.g., "tan#*trou*"). Pour ce faire, nous avons examiné les corrélations (par item) entre l'effet d'alignement sur les latences de *word spotting*, le taux de réponses correctes en catégorisation, et l'effet de la position de la frontière lexicale en répétition partielle, séparément pour les items VCCVC et CVCCV (voir Tableau 5). La logique sous-jacente consistait à exploiter les décisions d'ouverture et de clôture syllabique

comme révélateurs de la fonction perceptive des indices en question dans les deux tâches de segmentation lexicale. Par exemple, sous l'hypothèse que ces indices sont des indices de frontières de mots/syllabes, plutôt que des indices de débuts (ou de fins), la tâche de catégorisation devrait corréler de la même façon —pour ce qui est de l'effet de la position de la frontière— avec la tâche de Partie 1 (repérage des fins de syllabes) qu'avec celle de Partie 2 (repérage des attaques). Cette analyse permettait par ailleurs d'élucider les liens entre *word spotting* et catégorisation.

Tableau 5. Corrélations (r de Bravais-Pearson) entre l'effet d'alignement (C#C moins #CC) sur les TRs en *word spotting*, le taux de réponses correctes en catégorisation, et l'effet de la position de la frontière lexicale (C#C moins #CC) en répétition de première et de seconde partie, séparément pour les items VCCVC et CVCCV.

	Word spotting	Catégorisation	Partie 1	Partie 2
		Items VCCVC (n = 32)		
Word spotting	--	-.52 **	.24	.51 **
Catégorisation		--	-.12	-.44 *
Partie 1			--	.58 ***
Partie 2				--
		Items CVCCV (n = 32)		
Word spotting	--	-.19	.01	.28
Catégorisation		--	.18	-.36 *
Partie 1			--	.14
Partie 2				--

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

La performance de catégorisation, autrement dit la capacité des auditeurs à retrouver les intentions lexicales sous-jacentes, était corrélée négativement avec l'effet de la position de la frontière lexicale en répétition de Partie 2 ($r(32) = -.44$, $p < .02$, pour les items VCCVC; $r(32) = -.36$, $p < .05$, pour les items CVCCV), mais pas en répétition de Partie 1, quelle que soit la

structure phonologique des stimuli. Ceci suggère que les variations acoustico-phonétiques liées à l'assignation lexicale de l'obstruante sont perçues exclusivement comme des indices de début de mot/syllabe, et pas comme des indices de frontière.

De la même manière, l'effet d'alignement initial en *word spotting* (e.g., "ic#roche" vs. "i#croche") était corrélé significativement avec l'effet de la position de la frontière lexicale uniquement en répétition de Partie 2 ($r(32) = .51$, $p < .004$), pas en répétition de Partie 1. Ceci conduit donc à penser que les indices exploités pour localiser les débuts de mots potentiels ne sont pas perçus comme des indices de transition entre deux mots/syllabes, mais, encore une fois, uniquement comme des indices de début.

En outre, l'effet d'alignement final (e.g., "tant#rou" − "tan#trou") n'était corrélé ni avec l'effet de la position de la frontière lexicale en répétition de Partie 1, ni avec celui en répétition de Partie 2, ce qui colle parfaitement avec l'idée qu'il s'agit d'un effet indirect, probablement imputable à la présence d'un compétiteur (e.g., "trou") chevauchant, tantôt non-aligné, tantôt aligné sur l'attaque de la syllabe suivante (cf. la discussion de l'Expérience 1).

Enfin, l'effet d'alignement (en *word spotting*) était corrélé négativement avec la performance de catégorisation sur les items VCCVC, mais pas sur les items CVCCV. Ceci confirme l'idée que l'effet final (e.g., "tan#trou") est relativement indirect, et démontre du même coup que *word spotting* et catégorisation sont sensibles aux mêmes indices de début de mot (ou de syllabe). En somme, le patron de corrélations obtenu ici est exactement celui auquel on s'attendrait sur base de l'hypothèse qu'en tâche lexicale, les variations de durée mises en évidence dans l'Expérience 1 sont perçues comme des indices de début de mot (ou de syllabe), plutôt que comme des indices de frontière. En revanche, hormis la différence de rythme entre les deux types de structures phonologiques et le fait que la voyelle tronquée était initiale dans un cas et finale dans l'autre, il est difficile d'expliquer pourquoi l'effet de la frontière lexicale en répétition de Partie 1 n'était corrélé avec celui en répétition de Partie 2 que sur les items VCCVC ($r(32) = .58$, $p < .001$).

Discussion générale

Ce premier chapitre expérimental a examiné sur le front de la production, et sur celui de la perception de la parole, la présence d'indices sur la position des frontières de mots dans les groupes OBLI du français. Au regard des mesures des durées segmentales (cf. Expérience 1), il est clair que chez le locuteur particulier considéré, les énoncés du type …VC#CV…, tels que "tante rou(blarde)", diffèrent de leurs homologues du type …V#CCV…, tels que "temps trou(blant)", par un allongement substantiel de la voyelle qui précède la frontière, de l'obstruante de fin de mot et de la liquide initiale. La durée plus longue de la liquide initiale corrobore, sur les groupes consonantiques, l'allongement typique des consonnes intervocaliques et des voyelles à l'initiale de mots (ou de syllabes), par rapport à leurs variantes non-initiales (e.g., Barry, 1984; Lehiste, 1960; Oller, 1973, pour l'anglais; Dejean de la Bâtie, 1993; Vaissière, 1989; pour le français; Quené, 1992, 1993, pour le néerlandais; Kohler, 1976, pour l'allemand; Lehiste, 1964, pour le tchèque; Carlson *et al.*, 1972, pour le suédois; cf. chapitre 2, section 2.1.1.1, pour plus de détails).

En revanche, l'allongement de la voyelle précédente en position non-limite (…VC#CV) ne semble pas être un phénomène indépendant de la langue. Il semble même qu'il soit tout à fait spécifique à la situation d'enchaînement proprement dite en français. En anglais, les données sont totalement incohérentes: allongement, rétrécissement, ou encore absence d'effet laissent à penser que les situations expérimentales étaient peu contrôlées (Barry, 1981; Crystal & House, 1990; Hoard, 1966; Lehiste, 1960). En néerlandais, Quené (1992, 1993) a trouvé que le temps de retombée de l'énergie de la première voyelle était plus long en position limite qu'en position non-limite (e.g., "die pin" vs. "diep in"), mais il n'a trouvé aucune différence quant à la durée totale de cette voyelle. Plus intéressant encore, en français, les études de Dejean de la Bâtie (1993) et Spinelli *et al.* (2003) sur la liaison enchaînée (e.g., "petit ami" vs. "petit tamis"), situation pourtant très proche de l'enchaînement, n'ont trouvé aucune différence de durée fiable sur la voyelle précédente, en dépit d'un nombre important de locuteurs (cf. Spinelli *et al.*). Par contre, deux études phonétiques récentes sur l'enchaînement, initiées à la suite du présent travail, corroborent sur des séquences à singleton consonantique médian l'allongement de la voyelle précédente en position non-limite (Content *et al.*, 2004; Fougeron

et al., 2002). A l'aide de paires de séquences homophonémiques du type "cale égale" vs. "cas légal" lues en contexte de phrase, Fougeron *et al.* (2002) ont trouvé que la voyelle précédente était plus longue et avait un premier formant (F1) plus élevé lorsqu'elle était en position non-limite (VC#V) qu'en position limite (V#CV); par ailleurs, la consonne pivotale et la voyelle subséquente étaient plus longues en début de mot. Dans une étude complémentaire, Content *et al.* (2004) ont reproduit les résultats de Fougeron *et al.*, à l'aide de suites de pseudomots du type "nom + adjectif", telles que "grouk ifa" vs. "grou kifa". L'allongement de la voyelle précédente en position non-limite (VC#V) s'avère donc être un phénomène robuste et spécifique à l'enchaînement dans notre langue.

Etant donné ce qui vient d'être dit, le comportement de l'obstruante, plus longue en position finale (OB#LI) qu'initiale (#OBLI) semble aller à contre-courant. Peut-être s'agit-il d'un phénomène spécifique aux groupes consonantiques, voire spécifique aux groupes OBLI; peut-être s'agit-il d'une forme de marquage optionnel, voire stratégique, liée au fait que le locuteur n'était pas naïf quant aux objectifs de cette recherche. Nous évaluerons ces questions à la lumière des données du chapitre 5.

Indépendamment de celles-ci, du point de vue théorique, la mise en évidence de différences sur les durées segmentales (et sur le premier formant de la voyelle précédente) entre les conditions VC# et V#C, ici et dans les deux études complémentaires de Fougeron *et al.* (2002) et de Content *et al.* (2004), indique que le phénomène de resyllabation généralement prédit dans le cas de l'enchaînement en français n'a pas eu lieu, ou du moins pas complètement. Contrairement aux hypothèses tirées du modèle d'encodage phonologique proposé par Levelt et coll. (1989, 1992, 1999; Levelt & Wheeldon, 1994; Levelt *et al.*, 1999; voir aussi Cholin *et al.*, 2004), il est clair que dans le cas des groupes OBLI et des consonnes intervocaliques, la production d'une séquence VC#CV (ou VC#V) requiert la récupération ou la mise en branle de patrons articulatoires syllabiques différents de ceux impliqués dans la production d'une séquence homophonémique du type V#CCV (ou V#CV), où la consonne pivotale appartient à la représentation sous-jacente (ou lexème) du second mot. Le but de l'Expérience 6b (chapitre 5) sera de déterminer si la resyllabation est juste incomplète, ou bien

si elle est avant tout un *construct* nécessaire à la complétude de certaines théories phonologiques.

Sur le front de la perception, les données de *word spotting*, de catégorisation et de répétition partielle (Expériences 1-3) confirment indirectement la présence de co-variations phonétiques associées à la position de la frontière lexicale dans les groupes OBLI. Mais surtout, elles démontrent l'exploitation de ces indices dans la segmentation et la reconnaissance des mots, et leur influence sur la syllabation explicite. Au regard des corrélations entre la performance et les variations des durées segmentales, la voyelle précédente et la liquide s'avèrent être partie prenante dans la perception des intentions lexicales du locuteur. La détection de mots enchâssés en finale de porteurs était sensible à la durée de la liquide (e.g., "ic#roche" vs. "i#croche"); les choix de segmentation lexicale (en tâche de catégorisation) variaient selon les durées respectives de la liquide et de la voyelle précédente. Par ailleurs, sur base de l'analyse corrélationnelle entre les trois tâches, il semble que les corrélats de l'assignation lexicale de l'obstruante ne soient pas perçus comme des indices sur la position des frontières entre les mots (ou syllabes), c'est-à-dire les endroits du signal où un mot (ou une syllabe) finit et où un autre mot (ou une autre syllabe) débute, mais plutôt exclusivement comme des indices sur la localisation des débuts de mots (ou de syllabes). Autrement dit, la fonction de l'allongement de la voyelle précédente en position non-limite (…VC#CV…) ne serait pas d'informer l'auditeur que le mot (ou la syllabe) qui contient cette voyelle se termine avec l'obstruante, mais de lui signaler, comme le fait la liquide initiale (…VC#CV…), qu'un mot (ou une syllabe) débute sur la seconde consonne du groupe médian. Après tout, le traitement lexical ne requiert que de localiser les débuts de mots, le reste pouvant se faire indépendamment du signal.

Les résultats de l'expérience de catégorisation (e.g., "tante rou…" vs. "temps trou…") et les effets d'alignement initial (e.g., "ic#roche" vs. "i#croche") et final (e.g., tant#rou vs. tan#trou) en *word spotting* reproduisent, et corroborent, à l'aide d'une tâche *online* et de séquences de mêmes structures grammaticale et accentuelle, ceux obtenus par Quené (1992, 1993) en néerlandais (e.g., "diep in" vs. "die pin"). Ils corroborent aussi, à l'aide de deux paradigmes différents et d'un cas de resyllabation moins idiosyncratique, ceux de Spinelli *et al.* (2003) sur

la liaison. Au contraire du paradigme d'amorçage par répétition intermodale utilisé par ces auteurs, celui de *word spotting* (employé ici) n'implique *a priori* aucun recours à des représentations orthographiques. Par conséquent, l'obtention d'effets de l'alignement entre mots et structure phonétique en *word spotting* suggère fortement que les effets de Spinelli *et al.* reflètent l'influence directe des détails phonétiques sur l'activation des mots, plutôt que leur influence potentielle sur l'appariement pré-lexical entre sons et unités orthographiques (cf. Jacobs, Rey, Ziegler, & Grainger, 1998).

Pris séparément, les résultats de *word spotting* permettent également d'affiner l'interprétation de l'effet d'alignement initial entre mots et syllabes (e.g., "zun.lac" vs. "zu.glac") obtenu à l'aide du même paradigme par Dumay *et al.* (2002). Puisque les présents résultats montrent que les corrélats phonétiques de l'assignation lexico-syllabique de l'obstruante affectent la détection des mots enchâssés, notamment en finale de porteurs (e.g., "ic#roche" vs. "i#croche"), on peut dès lors penser que dans l'étude de Dumay *et al.*, l'effet d'alignement initial est imputable, au moins partiellement, aux détails phonétiques de la syllabation.

Enfin, la dernière expérience de ce chapitre montre que les choix de syllabation explicite en tâche de répétition de partie sont sensibles aux intentions lexicales du locuteur. Toutefois, contrairement aux expériences examinant l'influence de variables phonologiques ou orthographiques (Content *et al.*, 1999, 2001; Goslin *et al.*, 1999), ici, non seulement les décisions de clôture (e.g., "i]croche" vs. "ic]roche"), mais également celles d'ouverture syllabique (e.g., "i[croche" vs. "ic[roche"), étaient attirées vers la frontière lexicale sous-jacente. L'absence d'une dissociation en fonction de la partie à répéter (avec un effet uniquement en tâche de Partie 1) n'est sans doute pas surprenante étant donné la nature des indices manipulés. Cependant, on peut se demander si l'élément déterminant est l'activation différentielle des mots enchâssés, en fonction des différences phonétiques liées à l'assignation lexicale de l'obstruante qui détermine l'effet (cf. Davis *et al.*, 2002; Gow & Gordon, 1995; Spinelli *et al.*, 2003; Tabossi *et al.*, 1995); ou bien les indices phonétiques eux-mêmes, par leur participation à l'élaboration d'une représentation perceptive de la syllabation, laquelle recouvrirait partiellement la trace métaphonologique à la base du jugement (une manière de

dissocier ces deux interprétations serait d'utiliser des séquences sans signification, dont les portions situées de part et d'autre des deux points de segmentation potentiels ne correspondraient elles-mêmes à aucun mot).

Chapitre 5 Nature du groupe consonantique et indices acoustico-phonétiques des frontières de mots: comparaison entre les groupes obli et "/s/ + OBstruante"

Au chapitre précédent, nous avons mis en évidence à l'aide de la mesure des durées segmentales et de tests perceptifs, l'existence de variations acoustico-phonétiques codant la position des frontières de mots dans les groupes OBLI du français. Le but du présent chapitre est d'examiner le caractère général vs. spécifique de ces variations de bas niveau, en comparant les groupes OBLI à l'autre classe de groupes consonantiques potentiellement concernés par cette question: les groupes SOB (/s/ + OBstruante). Alors que les groupes OBLI sont systématiquement assignés en position d'attaque (i.e., /.OBLI/) par les diverses descriptions phonologiques de la syllabation du français, les groupes SOB font l'objet d'un traitement qui diffère selon la théorie ou l'algorithme utilisé: soit /.SOB/, soit /S.OB/ (voir Goslin & Frauenfelder, 2000, pour plus de détails; et aussi Bertinetto, 1999). De la même manière, les données sur la syllabation explicite des groupes SOB montrent beaucoup plus d'indécision, du moins de la part des locuteurs des langues romanes (Bertinetto, 2004), que celles relatives aux groupes OBLI, généralement assignés en attaque de syllabe, que ce soit dans les langues romanes ou germaniques (Berg, 2001; Bertinetto, 2004; Treiman *et al.*, 1992; cf. aussi les données de l'Expérience 3). Etant donné leurs traitements phonologiques différents, il se pourrait que ces deux classes consonantiques montrent également des divergences quant à leur manière d'encoder phonétiquement la position des frontières de mots.

La première expérience de ce chapitre a comparé les groupes OBLI et SOB dans l'effet de l'alignement final entre mot et structure phonétique (e.g., "tant#rou" vs. "tan#trou"; "ras#tu" vs. "ra#stu") en *word spotting*. La comparaison était restreinte à l'effet final, car il était impossible de trouver suffisamment de mots CCV débutant par /s/ et enchâssant un mot de structure CV. Au contraire du chapitre 4, les locuteurs ayant prêté leur voix étaient désormais naïfs quant aux objectifs de l'enregistrement. Ce changement de stratégie nous permettait

donc d'examiner la nature "automatique" des variations acoustiques, notamment sur les durées segmentales.

Expérience 4 – Effet d'alignement final en word spotting

Méthode

Participants. Quarante-huit étudiants (âge moyen: 20 ans, limites: 17-27) de l'Université libre de Bruxelles ont participé à l'expérience comme volontaires ou en échange de crédits dans le cadre d'un cours d'introduction à la psychologie expérimentale. Tous étaient de langue maternelle française, et ont rapporté ne souffrir d'aucun trouble de l'audition ou du langage. Aucun d'eux n'avait participé aux expériences précédentes.

Matériel. Trente-deux mots CVC ou CCVC (noms ou adjectifs; fréquence logarithmique moyenne: 3.39, limites: 1.66-5.77) étaient enchâssés comme cibles à l'initiale de séquences bisyllabiques sans signification, CVCCV ou CCVCCV (e.g., "race" dans "rastu"), et dont les groupes consonantiques médians étaient de type SOB (i.e., st, sp, sk et sf). La portion CV (ou CCV) des cibles était à dessein toujours un mot (e.g., "rat"). Comme dans l'Expérience 1, deux versions auditives de chaque porteur ont été extraites d'énoncés appariés de deux mots, différant quant à la position de la frontière lexicale: VC#CV (e.g., "ras#tu" extrait de *race tuméfiée*) ou V#CCV (e.g., "ra#stu" extrait de *rat stupéfait*). Ici, avec les groupes SOB, il était impossible que les portions finales CCV et CV soient appariées en termes de leur statut lexical. En l'occurrence, aucune finale CCV ne constituait un mot (e.g., "stu"), alors que c'était en général le cas de leur portion CV (e.g., "tu").

Pour la condition OBLI, 32 séquences semblables à celles de l'Expérience 1 ont été créées. Afin qu'elles soient similaires à celles de la condition SOB, aussi bien des mots CVC (e.g., "tante") que CCVC (e.g., "grappe") étaient utilisés comme cibles (fréquence logarithmique moyenne: 3.49, limites: 2.12-5.01). Par ailleurs, pour calquer l'absence de variation quant à l'identité de la première consonne dans les groupes SOB, l'identité de la liquide dans les groupes OBLI était toujours /ʀ/ (i.e., tʀ, gʀ, pʀ, dʀ, bʀ, vʀ, kʀ et fʀ).

Porteurs OBLI et SOB étaient appariés en termes de la fréquence logarithmique des cibles (3.49 vs. 3.39), de celle des compétiteurs enchâssés C(C)V (3.73 vs. 3.62), et de la fréquence logarithmique relative cible C(C)VC/compétiteur C(C)V (1.05 vs. 1.06; tous les *Fs* < 1). Hormis le fait que le locuteur (C.H.) était différent et qu'il ignorait tout de l'expérience, pour le reste, porteurs et remplisseurs —repris aux Appendices 4.1 et 4.2— étaient similaires à ceux de la condition d'enchâssement initial de l'Expérience 1.

Deux listes comprenant 64 porteurs et 64 remplisseurs en ordre pseudo-aléatoire fixe ont été composées. Toutes deux se scindaient en deux blocs équilibrés, dédiés respectivement à la détection des cibles CVC et CCVC. La position des cibles et des remplisseurs était identique d'une liste à l'autre. A l'intérieur des blocs, l'alignement et la nature du groupe consonantique variaient, la condition alignée, ou non-alignée, caractérisant une cible particulière étant contre-balancée à travers les listes. Dans un bloc, plus de trois porteurs ou trois remplisseurs n'apparaissaient jamais en succession directe. En outre, deux porteurs ou trois remplisseurs successifs n'appartenaient jamais à la même condition d'alignement. Une liste comportait dès lors 16 cibles par condition d'alignement pour les groupes OBLI, et 16 cibles par condition d'alignement pour les groupes SOB. A l'instar de la position d'enchâssement dans l'Expérience 1, chaque type de structure (CVC vs. CCVC) faisait l'objet d'un entraînement spécifique; celui-ci comprenait 30 essais, parmi lesquels 20 porteurs dont 15 en condition non-alignée.

Procédure. Vingt-quatre participants ont été testés avec chacune des listes. La structure des stimuli relevant de blocs distincts, l'ordre de présentation des blocs était contre-balancé à travers les participants. A l'exception de la période d'enregistrement du temps de réaction, qui était ici de 3 s, la procédure était identique à celle de la condition d'enchâssement initial de l'Expérience 1.

Résultats et discussion

Deux items (i.e., "fronde" et "plainte") omis par plus de 50% des participants ont été exclus des analyses. Les temps de réaction ont été ajustés en soustrayant la durée du CVC initial, afin d'être calculés à partir de la fin de la cible. Les analyses chronométriques ont été réalisées sur les temps de réaction accompagnés d'une réponse orale correcte. La distribution de ces temps

de réaction pour chaque condition est rapportée à l'Appendice 3 (Figure 10). Dans l'ensemble, le taux d'erreurs moyen (i.e., omissions et réponses manuelles sans réponse orale ou accompagnées d'une réponse orale incorrecte) s'élevait à 18% pour la condition OBLI et à 16.2% pour la condition SOB. Le taux de fausses alarmes était respectivement de 6.6% et 12.2%. Le Tableau 6 présente les temps de réaction moyens (par participant) et les taux d'erreurs moyens pour l'Expérience 4.

Tableau 6. Temps de réaction moyens (TRs) mesurés à partir de la fin de la cible, et taux d'erreurs moyens (Ers) pour l'Expérience 4. Les écarts-types figurent entre parenthèses.

OBLI	Type de Porteur	TRs (ms)	Ers (%)
Alignée	tant#rou	890 (226)	16.0 (10.6)
Non-alignée	tan#trou	974 (196)	20.0 (14.5)
Différence		84 *pi*	4 *p*
SOB			
Alignée	ras#tu	829 (186)	17.7 (12.4)
Non-alignée	ra#stu	837 (220)	14.7 (9.7)
Différence		8	-4

Note. "*pi*" indique une différence statistiquement significative par participant et par item, tandis que "*p*" indique une différence significative par participant seulement.

Les temps de réaction ont été soumis à des ANOVAs pour mesures répétées avec le groupe consonantique et l'alignement comme facteurs. L'effet du groupe consonantique était significatif par participant et frôlait le seuil de signification par item ($F_1(1,47) = 24.91$, $MSE = 18910.0$, $p < .001$; $F_2(1,60) = 3.88$, $MSE = 92385.9$, $p = .053$). En moyenne, les latences de détection étaient 99 ms plus longues dans la condition OBLI que dans la condition SOB (932 ms vs. 833 ms). L'alignement avait un effet significatif ($F_1(1,47) = 5.05$, $MSE = 19980.1$, $p < .03$; $F_2(1,60) = 7.51$, $MSE = 14188.0$, $p < .01$). Les cibles étaient détectées 46 ms plus tôt lorsqu'elles étaient alignées avec la frontière lexicale sous-jacente que lorsqu'elles ne l'étaient pas (859 ms vs. 905 ms). De façon remarquable, en dépit de la magnitude de l'effet

d'alignement dix fois plus grande dans les groupes OBLI que dans les groupes SOB, l'interaction entre l'alignement et le groupe consonantique n'était significative que par item ($F_1(1,47) = 2.81$, $MSE = 24269.5$, $p = .101$; $F_2(1,60) = 5.11$, $MSE = 14188.0$, $p < .03$). Comme l'indiquent des comparaisons planifiées menées sur les temps de réaction par item, l'effet de l'alignement était significatif sur les porteurs OBLI (107 ms (896 vs. 1003), $F_2(1,60) = 12.12$, $MSE = 14188.0$, $p < .001$), mais pas sur les porteurs SOB (10 ms (837 vs. 847); $F_2 < 1$).

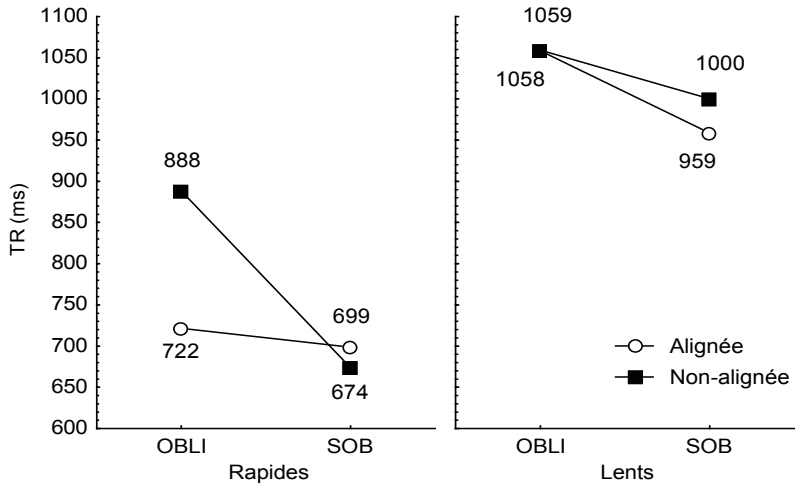

Figure 5. Temps de réaction moyens en fonction de l'alignement et du groupe consonantique pour les deux catégories de participants (rapides vs. lents).

Etant donné la présence d'une interaction uniquement par item, les données par participant ont été réanalysées en tenant compte de la rapidité de leurs réponses au moyen de la technique du partage par la médiane (cf. Dupoux, 1993; Fox, 1984; Miller & Dexter, 1988; Radeau & Morais, 1990; Sebastián-Gallés, Dupoux, Seguí, & Mehler, 1992). Les participants ont été répartis en deux catégories, "rapides" vs. "lents", sur base de leur temps de réaction moyen, et ce facteur binaire de rapidité a ensuite été rajouté dans l'ANOVA sur les latences de détection par participant (voir Figure 5). L'introduction du facteur "rapidité" ne modifiant fondamentalement pas les effets principaux des facteurs "nature du groupe consonantique" et "alignement", seuls l'effet et les interactions de cette variable avec les autres facteurs seront rapportés ici.

L'effet de la rapidité était hautement significatif (746 ms vs. 1019 ms; $F_1(1,46) = 105.13$, $MSE = 34141.5$, $p < .001$), et la rapidité n'interagissait ni avec l'alignement ($F_1(1,46) = 1.50$, $MSE = 19770.0$, $p > .2$) ni avec le groupe consonantique ($F_1 < 1$). En revanche, et ce point est fondamental, la rapidité modifiait de façon significative l'interaction groupe consonantique x alignement ($F_1(1,46) = 7.60$, $MSE = 31283.0$, $p < .01$). Pour décomposer cette interaction double, des ANOVAs ont dès lors été menées séparément pour les participants rapides et lents.

L'interaction alignement x groupe consonantique était significative chez les sujets rapides ($F_1(1,23) = 13.32$, $MSE = 16508.9$, $p < .002$). Comme l'attestent des comparaisons planifiées, l'effet de l'alignement était hautement significatif sur les porteurs OBLI (166 ms; $F_1(1,23) = 12.73$, $MSE = 26116.6$, $p < .002$), mais pas sur les porteurs SOB (-15 ms; $F_1(1,23) = 1.41$, $MSE = 5349$, $p > .2$). Enfin, l'effet du groupe consonantique était significatif dans la condition non-alignée (214 ms; $F_1(1,46) = 25.93$, $MSE = 21268.6$, $p < .002$), mais pas dans la condition alignée (23 ms; $F_1 < 1$). En revanche, chez les sujets lents, seul l'effet du groupe consonantique était significatif (80 ms; $F_1(1,23) = 7.01$, $MSE = 21641.8$, $p < .02$); il n'y avait aucun effet de l'alignement ($F_1 < 1$) ou interaction entre l'alignement et le groupe consonantique ($F_1 < 1$). La tâche de *word spotting* n'est donc ici sensible à l'alignement que chez les participants les plus rapides, mais elle démontre que la présence d'un effet de l'alignement entre mots et intentions lexicales sur les latences de détection est dépendante de la nature du groupe consonantique médian[22].

Les taux d'erreurs ont été soumis à des ANOVAs similaires à celles effectuées sur les temps de réaction. Ici, seule l'interaction alignement x groupe consonantique était significative ($F_1(1,47) = 8.64$, $MSE = 68.5$, $p < .01$; $F_2(1,60) = 4.27$, $MSE = 89.3$, $p < .05$; groupe consonantique: $F_1(1,47) = 1.43$, $MSE = 105.7$, $p > .2$; tous les autres $Fs < 1$), la différence entre les deux conditions d'alignement s'inversant d'une condition consonantique à l'autre. Comme en témoignent des comparaisons planifiées, l'effet de l'alignement (4%) sur les

[22] Bien que, comme le soutiennent Floccia, Kolinsky, Dodane et Morais (2003), le *word spotting* pourrait avoir une composante métaphonologique, l'obtention d'effets acoustico-phonétiques tels que ceux-ci, qui plus est sur les temps de réaction les plus courts, suggère que la tâche est néanmoins sensible aux niveaux les plus précoces du traitement de la parole.

groupes OBLI était significatif par participant et marginalement significatif par item ($F_1(1,47) = 4.54$, $MSE = 85.8$, $p < .04$; $F_2(1,60) = 2.72$, $MSE = 89.3$, $p = .10$). En revanche, la tendance opposée (-4%) sur les porteurs SOB, laquelle suggérait la présence d'un léger échange entre rapidité et précision, n'était que marginalement significative par participant ($F_1(1,47) = 3.64$, $MSE = 59.0$, $p = .062$; $F_2(1,60) = 1.61$, $MSE = 89.3$, $p > .2$). Dans l'ensemble, le patron d'erreurs corroborait donc celui observé sur les temps de réaction.

Tableau 7. <u>En haut:</u> Durées segmentales moyennes (ms) en fonction de la nature du groupe consonantique et de la position de la frontière lexicale (#), telles qu'évaluées par l'auteur (N.D.) et par le juge indépendant (M.C.). <u>En bas:</u> Valeurs du coefficient de corrélation (r de Bravais-Pearson) entre les durées segmentales rapportées par l'auteur et celles rapportées par le juge indépendant.

C_2C_3	#	C_1 (+C_{1b})	V_1	C_2	C_3	V_2
<u>N.D.</u>						
OB + /ʀ/	C#C	134 (53)	125 (38) ***	73 (18)	78 (15) ***	59 (15)
	#CC	131 (50)	104 (24)	76 (19)	53 (18)	59 (19)
/s/ + OB	C#C	131 (53)	96 (31)	99 (28)	85 (16)	55 (12)
	#CC	134 (53)	92 (25)	99 (24)	79 (14)	59 (19)
<u>M.C.</u>						
OB + /ʀ/	C#C	134 (51)	125 (39) ***	68 (23)	85 (20) ***	59 (13)
	#CC	131 (49)	103 (26)	67 (21)	62 (21)	61 (19)
/s/ + OB	C#C	127 (51)	94 (33)	103 (27)	90 (17)	53 (13)
	#CC	131 (52)	89 (25)	106 (22)	83 (15)	58 (17)
		<u>Valeurs du coefficient de corrélation r de Bravais-Pearson:</u>				
OB + /ʀ/	C#C	.99	.96	.85	.69	.86
	#CC	.99	.83	70	.84	.93
/s/ + OB	C#C	.98	.93	.94	.90	.94
	#CC	.98	.93	.93	.88	.84

<u>Note.</u> Une cellule représente 32 données. C_{1b} désigne la seconde consonne (glide ou liquide) du groupe d'attaque des 32 cibles CCVC. Les durées des segments affectés par la procédure d'extraction des stimuli sont rapportées en gris. Les différences significatives en fonction de la position de la frontière lexicale sont notées "*" si $p < .05$, "**" si $p < .01$, et "***" si $p < .001$.

Afin d'établir la source de l'interaction entre l'effet de l'alignement et celui du groupe consonantique, des mesures des durées segmentales ont été effectuées comme précédemment sur chaque porteur. Ici, cependant, la durée de chaque segment a été mesurée deux fois: une fois par l'auteur, et une fois par un juge indépendant, expérimenté dans la préparation des stimuli de parole. Celui-ci était totalement naïf quant aux tenants et aboutissants des mesures qu'il avait à effectuer. Le double mesurage permettait d'une part, de vérifier la fiabilité des durées acoustiques rapportées par l'auteur, et d'autre part de tester la cohérence entre les juges. Le Tableau 7 rapporte les durées segmentales moyennes selon l'auteur et le juge indépendant, et les coefficients de corrélation entre les mesures de l'un et celles de l'autre.

Cohérence entre les juges. Les durées moyennes obtenues par N.D. et M.C. étaient très proches les unes des autres, voire identiques dans certains cas, la différence la plus importante (de seulement 9 ms) se situant au niveau des groupes OBLI, où N.D. plaçait la fin de l'obstruante plus tardivement que M.C.. Comme l'indiquent les valeurs élevées des coefficients de corrélation (de .89 en moyenne), le patron de variations rapporté par M.C. concordait de façon remarquable avec celui rapporté par N.D.. Notons que même pour la mesure toujours délicate de /ʀ/ (voir Autesserre & Chafcouloff, 1998), la liquide des groupes OBLI, les corrélations obtenues (de .69 et .84) étaient plus que satisfaisantes. De manière à corroborer cette analyse, des ANOVAs, avec le juge, la position de la frontière lexicale et le groupe consonantique comme facteurs, ont été menées séparément pour chaque segment. Quel que soit le segment examiné, ni l'interaction de premier ordre juge x position de la frontière, ni l'interaction de second ordre (juge x groupe consonantique x position de la frontière) n'était significative (tous les $Fs < 1$, sauf pour l'interaction de second ordre, marginalement significative sur C_2: $F(1,62) = 2.87$, $MSE = 56.2$, $p = .096$). Corrélations et ANOVAs démontraient donc que les deux juges étaient sensibles avec une acuité équivalente aux mêmes éléments du paysage "oscillo-sonore".

Variations sur les durées segmentales. Les durées segmentales obtenues par N.D. et M.C. ont fait l'objet d'ANOVAs menées séparément pour chaque segment, en fonction de la nature du groupe consonantique et de la position de la frontière lexicale. Etant donné la similitude entre les deux patrons de résultats, seules les ANOVAs relatives aux mesures effectuées par

N.D. seront rapportées ici. L'interaction groupe consonantique x position de la frontière était significative pour V_1 ($F(1,62) = 7.17$, $MSE = 339.1$, $p < .01$) et C_3 ($F(1,62) = 18.96$, $MSE = 153.9$, $p < .001$). Comme en témoignent des comparaisons planifiées, V_1 était significativement plus longue (de 20.8%) en position non-limite qu'en position limite dans le cas des groupes OBLI ($F(1,62) = 20.69$, $MSE = 339.1$, $p < .001$), mais pas dans le cas des groupes SOB (5%; $F < 1$). Il en était presque de même pour C_3, qui était significativement plus longue (de 41.8%) à l'initiale qu'en finale de mot dans le cas des groupes OBLI ($F(1,62) = 65.32$, $MSE = 153.9$, $p < .001$), et montrait une tendance marginale à l'allongement initial (de 8%) dans le cas des groupes SOB ($F(1,62) = 3.70$, $MSE = 153.9$, $p = .06$), mais rien en comparaison des OBLI. En revanche, sur les autres segments, ni l'effet de la position de la frontière, ni l'interaction groupe consonantique x position de la frontière n'étaient significatifs (tous les *Fs* < ou proches de 1). En somme, le locuteur naïf C.H. reproduisait sur les groupes OBLI les deux allongements les plus importants observés chez le locuteur non-naïf des expériences précédentes (28.8% pour V_1; 42.9% pour la liquide), mais ne montrait aucun phénomène de durée équivalent et/ou fiable sur les groupes SOB.[23]

Pris ensemble, les résultats de l'expérience de *word spotting* et les mesures des durées segmentales suggèrent fortement que la présence d'indices phonétiques codant la position des frontières lexicales dans les groupes consonantiques du français dépend crucialement de la nature du groupe consonantique. En accord avec l'idée que l'enchaînement ne conduit pas à une resyllabation totale, les groupes OBLI montrent des variations des durées segmentales selon l'assignation lexicale de l'obstruante, et induisent des effets d'alignement dans une tâche de segmentation lexicale *online*. En revanche, les groupes SOB ne montrent aucune variation équivalente sur les durées segmentales et n'engendrent aucun effet perceptif. Selon ce point de vue, au contraire des alternatives à groupe OBLI, celles à groupe SOB seraient donc homophones au sens strict du terme.

[23] Tout comme dans l'Expérience 1, il n'y avait dans l'Expérience 4 aucune corrélation négative significative entre les différences de durée de cible (alignée moins non-alignée), et les effets d'alignement en *word spotting* (OBLI: $r(32) = .05$, $p > .7$ sur les TRs; $r(32) = .27$, $p > .1$ sur les Ers; SOB: $r(32) = -.14$, $p > .4$ sur les TRs; $r(32) = -.10$, $p > .5$ sur les Ers).

Toutefois, avant de pouvoir affirmer cela, il importe de considérer une différence importante entre les porteurs à groupe OBLI et ceux à groupe SOB utilisés dans l'Expérience 4. Dans les porteurs OBLI (e.g., "tantrou"), la portion finale CCV était toujours un mot ("trou"), et donc un compétiteur chevauchant de la cible initiale. En revanche, pour des raisons liées au lexique français, la portion équivalente (e.g., "stu") dans les porteurs SOB ("rastu") n'était jamais un mot. Or, si ce qui est important pour la segmentation lexicale est avant tout de localiser les débuts de mots —comme le suggère notamment l'hypothèse SOSH (Content *et al.*, 2001; Dumay *et al.*, 2002)— une manière d'interpréter l'effet d'alignement final sur les porteurs OBLI consiste à dire que celui-ci n'émerge qu'en raison de la présence d'un compétiteur débutant sur la consonne pivotale (cf. la discussion de l'Expérience 1). A l'appui de l'idée que les cibles OBLI (e.g., "tante") subissaient l'influence néfaste du compétiteur chevauchant (e.g., "trou"; cf. Vroomen & de Gelder, 1995), on remarquera d'ailleurs la différence de plus de 130 ms en faveur des latences de détection sur les porteurs SOB, en dépit du fait que les conditions OBLI et SOB étaient appariées en termes de la fréquence des cibles (e.g., "tante"; "race"; cf. Freedman, 1992) et de la fréquence relative entre les cibles et les compétiteurs enchâssés (e.g., "temps"; "rat"). Par conséquent, indépendamment des mesures des durées segmentales, on peut se poser la question suivante: l'absence d'effet de l'alignement entre mot et structure phonétique sur les porteurs SOB est-elle liée à l'absence de compétiteur CCV pour le segment /s/, ou est-elle le reflet d'une absence véritable de différences acoustico-phonétiques entre les alternatives C#C et #CC? En effet, bien que les mesures des durées segmentales ne montrent pour les groupes SOB aucune différence équivalente à celles observées sur les groupes OBLI, on ne peut exclure la possibilité que l'assignation lexicale du segment /s/ des porteurs SOB se marque par d'autres procédés phonétiques. On sait, par exemple, que dans le cas de l'assignation lexicale d'une consonne pivotale intervocalique, des changements de fréquences formantiques sur la voyelle précédente côtoient les phénomènes de durée (cf. Content *et al.*, 2004).

Par ailleurs, une autre question que l'on peut se poser est celle de la "généralisation perceptive". Bien qu'en moyenne les participants ne montrent aucun effet de l'alignement sur les porteurs SOB, on peut se demander s'ils exploitent néanmoins les différences de durée les

plus marquées, ou s'ils y restent sourds. En faveur de la première possibilité, il y avait, dans la condition SOB, une corrélation négative marginale entre l'allongement de l'obstruante et l'effet (non-significatif) de l'alignement sur les latences de détection ($r(32)= -.34$, $p = .056$; V1: $r(32)= -.29$, $p > .1$; /s/: $r(32)= .24$, $p > .15$; OBLI: tous les $rs(32)$ entre -.04 et .13, $ps > .4$). Cette corrélation laisse donc supposer que les auditeurs pourraient être sensibles à certaines différences de durée dans les groupes SOB. La question de l'homophonie des alternatives à groupe SOB, et celle de la généralisation perceptive constituaient les deux motivations principales de l'expérience suivante.

Expérience 5 – Perception des frontières de mots en tâche d'appariement co-déterminé

Le but premier de cette expérience était de déterminer si les séquences alternatives, C#C et #CC, à groupe SOB sont véritablement homophones, la situation contrôle étant constituée des séquences à groupe OBLI. La tâche des participants, similaire à celle employée par Lehiste (1960), était dite d'"appariement co-déterminé". Les participants devaient déterminer, parmi deux séquences auditives successives (e.g., "ras#tu" – "ra#stu"), celle qui correspondait le mieux à l'énoncé spécifié par écrit (e.g., "rat stupéfait"). Au contraire de la tâche de catégorisation (cf. Expérience 2), l'appariement co-déterminé donne la possibilité aux participants de comparer les deux versions acoustiques d'une même séquence au sein d'un même essai. Cette situation leur permet donc de baser leur choix de réponses non seulement sur les propriétés intrinsèques à chaque séquence, mais aussi sur les différences d'impression auditive entre les deux alternatives. A supposer que la performance d'appariement soit supérieure au niveau du hasard (50%) dans la condition OBLI et qu'elle n'en diffère pas dans la condition SOB, on conclura que les alternatives à groupe SOB sont homophones. Puisque l'auteur avait l'oreille particulièrement familiarisée avec les stimuli, il a lui aussi été soumis à la tâche.

Méthode

Participants. Outre la personne ayant conçu les stimuli, 24 étudiants (âge moyen: 21 ans, limites: 18-28) de l'Université libre de Bruxelles ont été testés en échange de crédits dans le cadre d'un cours de psychologie expérimentale. Tous étaient de langue maternelle française et rapportaient ne souffrir d'aucun trouble de l'audition ou du langage. Aucun d'eux n'avait participé aux expériences précédentes.

Matériel. Les séquences de parole étaient les 64 paires de séquences porteuses utilisées dans l'Expérience 4: 32 à groupe OBLI (e.g., "tant#rou" – "tan#trou") et 32 à groupe SOB (e.g., "ras#tu" – "ra#stu"). Deux listes auditives ont été créées. Elles comprenaient chacune les 64 paires de séquences mélangées en ordre pseudo-aléatoire. Alors que la position d'une paire particulière ne variait pas d'une liste à l'autre, l'ordre de présentation des séquences était contre-balancé à travers les listes. Celles-ci ne contenaient en succession directe jamais plus de trois paires de séquences de la même classe consonantique (OBLI vs. SOB). Elles se scindaient en deux blocs équilibrés, dédiés respectivement aux porteurs CVCCV (e.g., "rastu") et CCVCCV (e.g., "plascan"). En miroir des listes auditives, deux listes visuelles ont été composées. Elles contenaient chacune 32 énoncés du type C#C (e.g., "race tuméfiée") et 32 du type #CC (e.g., "plat scandinave"). A travers les listes, la position de la frontière lexicale était contre-balancée.

Procédure. Les participants ont été testés individuellement dans une pièce isolée. Les séquences auditives étaient présentées à un volume d'écoute confortable, au moyen d'un lecteur DAT Sony DTC-60ES et d'écouteurs Beyerdynamic DT-770. Les énoncés d'origine étaient quant à eux présentés par écrit en une colonne centrée sur les feuillets d'un protocole. La tâche des participants était de déterminer, parmi les deux séquences auditives présentées, celle qui d'après eux provenait de l'énoncé spécifié par écrit.

Chaque essai débutait par un signal de 100 ms. A ce signal, les participants avaient pour tâche d'articuler plusieurs fois l'énoncé visuel en étant particulièrement attentifs à la façon dont ils en enchaînaient les deux termes. Au bout de 8 s, un autre signal prévenait les participants de l'imminence des séquences auditives, lesquelles étaient jouées une demi-seconde plus tard. La paire de séquences était présentée trois fois, avec un intervalle de 300

ms entre les deux séquences, et un autre de 1500 ms entre deux présentations d'une même paire. Au terme de la troisième présentation, les participants disposaient de 8 s pour faire impérativement leur choix. Ils répondaient en entourant le chiffre "1" ou le chiffre "2" dont était flanqué l'énoncé. Revenir à un essai antérieur était interdit.

Douze étudiants ont été assignés à chaque liste auditive combinée, pour six d'entre eux, avec une liste visuelle, et pour les six autres, avec l'autre liste visuelle. Comme une liste se scindait en deux blocs équilibrés, l'ordre de présentation des blocs était contre-balancé à travers les participants. L'auteur (N.D.) a été assigné aléatoirement à l'une des quatre combinaisons d'une liste auditive avec une liste visuelle. L'expérience débutait par un bloc d'entraînement de huit essais. Elle durait environ 40 minutes.

Résultats et discussion

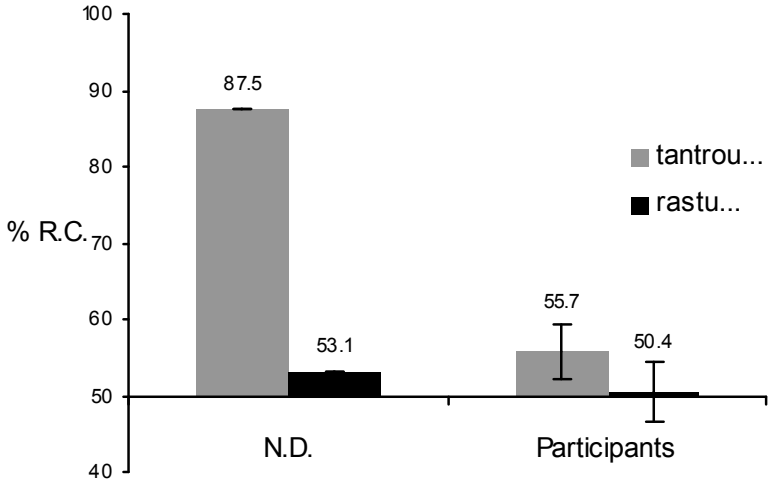

Figure 6. Taux (ou taux moyen) de réponses correctes (et intervalle de confiance) en fonction du groupe consonantique pour N.D. et les 24 participants naïfs.

Les taux bruts (ou taux moyens) de réponses correctes pour les conditions OBLI et SOB sont présentés à la Figure 6, séparément pour N.D. et les 24 participants naïfs. Sur le plan descriptif, les participants naïfs avaient un score d'appariement meilleur dans la condition OBLI que dans la condition SOB (55.7% (8.8) vs. 50.4% (9.3)). Mais, trop faible (i.e., 5.3%), cette différence ne ressortait que de façon marginalement significative à des tests-t bilatéraux effectués par participant et par item ($t_1(23) = 1.90$, $p = .072$; $t_2(62) = 1.73$, $p = .088$). Comme

en témoignent des tests-t unilatéraux, la performance des participants dans la condition OBLI, bien que peu élevée, était significativement supérieure au niveau du hasard ($t_1(23) = 3.20$; $p < .002$; $t_2(31) = 2.40$; $p < .02$). En revanche, leur performance dans la condition SOB, sensiblement identique au niveau du hasard, n'en était pas statistiquement différente ($t_1(23) = .21$, ns; $t_2(31) = .20$, ns).

Un test Khi-carré démontrait que la performance de N.D. était significativement modulée par la nature du groupe consonantique médian ($\chi^2(1, n = 64) = 9.06$, $p < .01$). Cette différence, d'ailleurs massive, allait dans le même sens que celle observée chez les participants naïfs (OBLI: 87.5% vs. SOB: 53.1%). En outre, comme des tests binomiaux en témoignent, N.D. avait une performance supérieure au niveau du hasard dans la condition OBLI ($z(n = 32) = 4.07$, $p < .001$), mais pas dans la condition SOB ($z(n = 32) = .18$, $p > .4$). Les résultats de N.D. confirmaient donc ceux des participants naïfs d'une manière beaucoup plus nette.

Au regard de ces résultats, et en dépit de l'extrême difficulté de la tâche pour les participants naïfs, il est à présent certain que les alternatives à groupe consonantique SOB qui ne varient que par l'assignation lexicale du segment /s/ (e.g., ras#tu… vs. ra#stu…) sont des homophones au sens le plus pur du terme. Même pour N.D., pourtant manifestement très entraîné à la détection de ce qui s'avère être de subtils indices de segmentation, seules les séquences à groupe OBLI autorisaient la récupération de la segmentation sous-jacente. Sur base de cette dissociation, on peut donc déduire que l'absence d'effet de l'alignement sur les séquences à groupe SOB dans l'expérience de *word spotting* est avant tout imputable à l'absence de variations codant la position de la frontière lexicale; la présence d'un compétiteur chevauchant, révélateur potentiel du degré d'alignement entre la cible et la structure phonétique, n'aurait probablement rien changé à la donne.

En outre, bien qu'il n'y ait ici aucune corrélation significative entre les différences de durée sur la voyelle V_1 et les deux consonnes médianes, et la performance d'appariement (tous les $r_s(32)$ entre -.05 et .24, *ps* > .15, pour les OBLI; tous les $r_s(32)$ entre -.29 et .04, *ps* > .1, pour les SOB), la co-variation des scores d'appariement moyens avec la présence de variations des durées segmentales suggère elle aussi, à sa manière, que les mécanismes de segmentation exploitent les durées segmentales. Par ailleurs, en accord avec l'idée que la tâche

d'appariement était influencée par les mêmes facteurs phonétiques que le *word spotting*, la corrélation négative entre le taux de réponses correctes par item (pour les participants naïfs) dans la première tâche et l'effet d'alignement (i.e., C#C moins #CC) sur les latences de détection dans la seconde était significative pour les groupes OBLI ($r(30) = -.46$, $p < .02$), mais pratiquement nulle pour les groupes SOB ($r(32) = -.003$, $p > .9$).

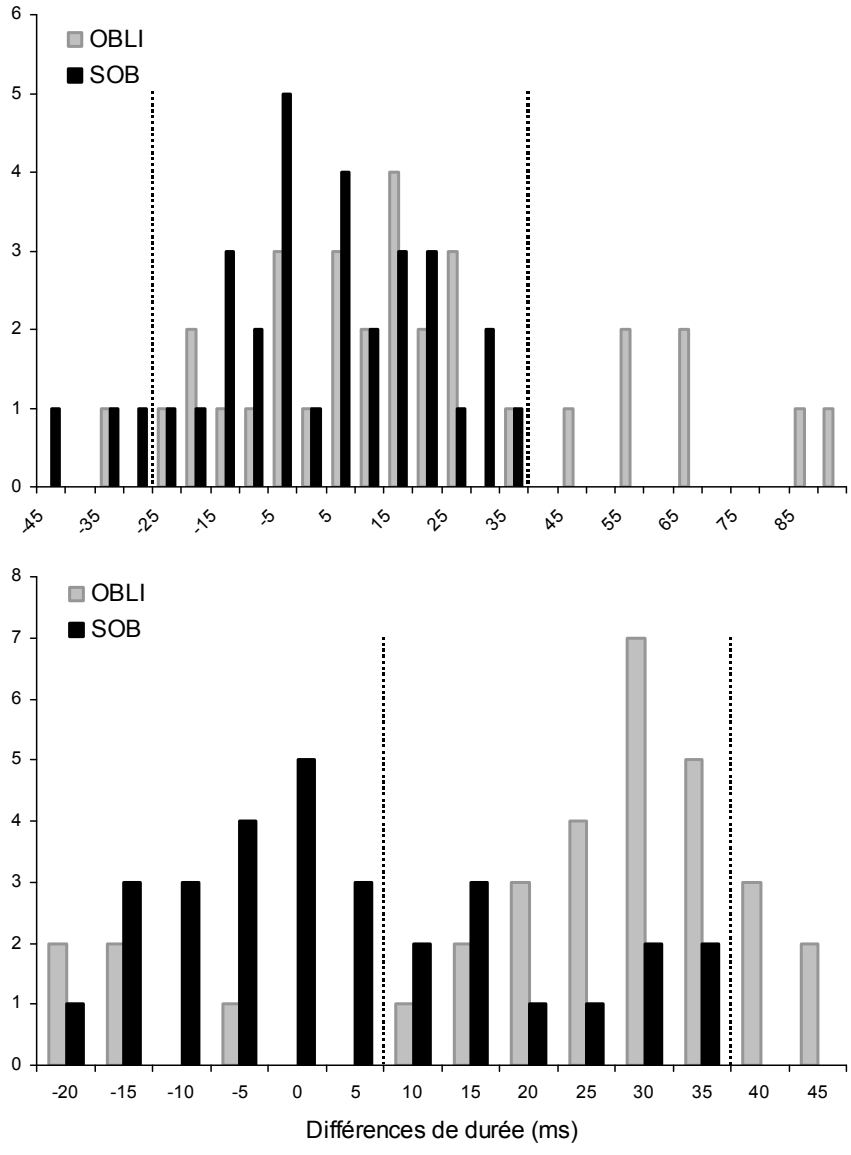

Figure 7. Distributions des différences de durée entre la condition C#C et la condition #CC pour V_1 (en haut) et C_3 (en bas).

Enfin, pour examiner la question de la "généralisation perceptive" abordée dans la discussion de l'Expérience 4, nous avons calculé les distributions des différences de durée sur la voyelle précédente (V_1) et la seconde consonne médiane (C_3), séparément pour les séquences OBLI et les séquences SOB. L'objectif de cette analyse était d'évaluer la performance d'appariement uniquement pour les paires de séquences OBLI et SOB montrant des variations de durée similaires sur les segments V_1 et C_3. Si la performance d'appariement pour la condition SOB de ce sous-ensemble d'items est supérieure au niveau du hasard et similaire à celle de la condition OBLI, on conclura que les variations des durées segmentales, pour peu qu'elles soient d'une magnitude suffisante, fonctionnent comme des indices de segmentation quelle que soit la nature du groupe consonantique. En revanche, si le patron de performance reste inchangé, on conclura que les différences de durée pourtant de même force sur les porteurs SOB ne sont pas exploitées.

Comme on peut le voir à la Figure 7, la distribution SOB chevauchait la distribution OBLI par les valeurs de différences situées entre -25 et 40 dans le cas de V_1 (cf. le haut de la figure), et 10 et 40 dans le cas de C_3 (cf. le bas). L'intersection entre ces deux zones de chevauchement impliquait 14 paires OBLI et 11 paires SOB. Dans l'ensemble, le patron de résultats restreints était similaire sur le plan descriptif à celui obtenu avec la totalité des paires. La performance des participants naïfs était de 53.9(11.7)% dans la condition OBLI et de 49.2(13.9)% dans la condition SOB. Bien que cette différence n'était pas fiable sur le plan statistique ($t_1(23) = 1.11$, $p > .2$; $t_2(23) = .99$, $p > .3$), le taux d'appariement était toutefois marginalement supérieur au niveau du hasard dans la condition OBLI ($t_1(23) = 1.61$; $p = .06$; $t_2(13) = 1.38$; $p = .095$), et n'en était pas statistiquement différent dans la condition SOB ($t_1(23) = .27$; $p > .3$; $t_2(10) = .20$; $p > .4$). De la même manière, la performance de N.D. était en moyenne plus élevée dans la condition OBLI (85.7%) que dans la condition SOB (72.7%). Bien que cette différence n'était pas significative ($\chi^2(1,24) = 24.73$, $p > .4$), la performance de l'auteur était différente du niveau du hasard dans la condition OBLI ($z(n = 14) = 2.41$, $p < .01$), mais pas dans la condition SOB ($z(n = 11) = 1.21$, $p > .1$). En somme, bien qu'en raison du nombre restreint d'items, ces résultats soient assez faibles sur le plan statistique, ils suggèrent fortement que l'exploitation des variations des durées segmentales comme indices de

segmentation est conditionnée par la nature des segments concernés, et n'est donc pas généralisable. Si les séquences SOB montrent occasionnellement des différences de durée selon l'assignation lexicale du /s/ équivalentes à celles des séquences OBLI, le système perceptif est tel qu'il y reste sourd, ou du moins est incapable de les utiliser afin de localiser les frontières de mots.

Si cette conclusion est vraie, alors il y a de fortes chances que l'absence de variations cohérentes sur les durées segmentales des séquences à groupe SOB ne soit pas spécifique au locuteur C.H., mais plutôt un phénomène indépendant, reflétant le fonctionnement du système phonologique du français. Le but de l'expérience suivante était de vérifier la nature générale du patron de variations mis en évidence chez C.H., à l'aide d'un large corpus produit par huit autres locuteurs naïfs. Le matériel consistait en un sous-ensemble des énoncés utilisés dans les Expériences 1 et 4. Les locuteurs avaient à les produire non seulement isolément, mais aussi en contexte de phrase.

Expérience 6 – Production d'énoncés OBLI et SOB isolément et en contexte de phrase

Méthode

Participants. Quatre femmes et quatre hommes de langue maternelle française ont été testés. Aucun d'eux ne souffrait de troubles de l'audition ou du langage et ne connaissait les tenants ou les aboutissants de l'expérience.

Matériel. Le matériel —repris à l'Appendice 5.1— était constitué de 48 paires d'énoncés de deux mots (CVC#CV… – CV#CCV…), telles que "*tante roublarde*" vs. "*temps troublant*". Parmi elles, 32 étaient à groupe consonantique OBLI, dont 16 avec la liquide /l/ et 16 avec la liquide /ʀ/, utilisées dans l'Expérience 1, et 16 étaient à groupe SOB utilisées dans l'Expérience 4. Dans le cas des groupes OBLI, l'obstruante était voisée dans la moitié des cas, et de manière équiprobable pour les deux liquides.

Les 96 énoncés étaient mélangés aléatoirement en deux blocs équilibrés. D'un bloc à l'autre, les énoncés d'une même paire occupaient des places identiques, tandis que la position

de la frontière lexicale était contre-balancée. La réalisation isolée des énoncés, décrits au locuteur comme des groupes de sens, et non de simples paires de mots, consistait en une lecture enchaînée des deux termes. Pour la réalisation en contexte de phrase, les énoncés étaient insérés immédiatement après le verbe en tant que complément de celui-ci dans des phrases appariées en termes du nombre de syllabes et de la position sérielle (comptée en syllabes) de l'énoncé. Comme l'indiquaient des ANOVAs, les trois conditions consonantiques ("OB + /l/", "OB + /ʀ/" et "/s/ + OB") ne différaient pas statistiquement, que ce soit en termes du nombre de syllabes par phrase (13.9 (2.6) vs. 12.4 (1.9) vs. 12.4 (2.6); $F(2,45) = 2.10$, $MSE = 5.5$, $p > .1$) ou de la position syllabique de l'énoncé (6.6 (1.9) vs. 6.8 (1.1) vs. 6.7 (1.5); $F < 1$). Dans 50% des cas, le reste de la phrase (e.g., "du brabant" et "d'avant-guerre") se rapportait au premier terme du complément, comme dans "*Aurélie est une tante roublarde du Brabant*", appariée avec "*Il naquit en ces temps troublants d'avant-guerre*"; pour les autres, il se rapportait au verbe, comme dans "*J'ai souffert d'une mince cataracte comme un martyre*", appariée avec "*Il nettoie la main scarifiée à l'aide d'éther*".

Procédure. Les enregistrements ont eu lieu dans une pièce isolée. Les participants avaient pour consigne de lire chaque phrase silencieusement, puis de la communiquer à l'expérimentateur. Les productions verbales étaient capturées à une fréquence d'échantillonnage de 44.1 kHz et un taux de conversion de 16 bits à l'aide d'un microphone Shure-12A et d'un enregistreur DAT Sony DTC-60ES. Enoncés de deux mots puis phrases étaient présentés en une liste centrée sur les feuillets d'un protocole. Comme pour les enregistrements précédents, chaque feuillet débutait par un item supplémentaire non-pertinent et se terminait par deux. Cette précautions avait pour but d'éviter que les productions verbales pertinentes ne soient affectées par les corrélats prosodiques de début ou de fin de liste. Les huit locuteurs ont tous été confrontés aux deux tâches, en commençant systématiquement par la production isolée. Chaque tâche faisant l'objet de deux blocs équilibrés, l'ordre de présentation des blocs était contre-balancé à travers les locuteurs. Une fois enregistrées, les portions pertinentes (CVCCV) des énoncés étaient transférées sur le disque dur d'un Macintosh Quadra-700 en gardant la même fréquence d'échantillonnage et le même taux de

conversion. La durée des segments phonémiques de chaque séquence —soit un total de 7680 phonèmes— a ensuite été mesurée par l'auteur de la même manière que précédemment.

Résultats et discussion

Tableau 8. Durées segmentales moyennes (ms) en fonction du contexte de réalisation, du groupe consonantique et de la position de la frontière lexicale (#). Les écarts-types figurent entre parenthèses.

Réalisation	C_2C_3	#	C_1	V_1	C_2	C_3	V_2
Isolée	OB + /l/	C#C	106 (11)	143 (20) *pi*	89 (18) *pi*	61 (8) *pi*	101 (10)
		#CC	105 (16)	115 (20)	83 (12)	47 (4)	99 (11)
	OB + /ʀ/	C#C	88 (12)	135 (18) *pi*	63 (12) *pi*	99 (19) *pi*	85 (6)
		#CC	91 (14)	111 (20)	59 (10)	82 (14)	79 (5)
	/s/ + OB	C#C	95 (10)	119 (13)	95 (17) *pi*	88 (8) *pi*	73 (9)
		#CC	96 (12)	116 (17)	89 (16)	83 (7)	83 (7)
En phrase	OB + /l/	C#C	68 (7)	122 (15) *pi*	68 (10)	49 (7) *pi*	97 (12)
		#CC	66 (8)	102 (13)	67 (8)	41 (3)	94 (12)
	OB + /ʀ/	C#C	59 (8)	108 (12) *pi*	48 (10)	87 (14) *pi*	81 (10)
		#CC	56 (4)	89 (11)	50 (7)	74 (11)	73 (6)
	/s/ + OB	C#C	61 (8)	101 (10)	73 (13)	81 (5) *pi*	67 (11)
		#CC	63 (6)	96 (10)	77 (13)	77 (5)	74 (12)

Note. Une cellule représente 128 données. La durée de V_2 (en gris) n'est rapportée qu'à titre indicatif, la consonne subséquente n'étant pas la même dans la condition C#C que dans la condition #CC. "*pi*" indique une différence statistiquement significative par participant et item, tandis que "*p*" indique une différence significative par participant seulement.

Les durées segmentales moyennes par participant et par condition sont présentées au Tableau 8. Chaque position phonémique a fait l'objet d'ANOVAs par participant et item, avec le contexte de réalisation, le groupe consonantique et la position de la frontière lexicale comme facteurs. Les variations de durée de la seconde voyelle (V_2) étant nécessairement contaminées par le fait que la consonne subséquente n'avait pas la même identité d'un énoncé à l'autre, les résultats qui les concernent ne seront pas rapportés. Pour les séquences OBLI, les allongements caractéristiques de la voyelle précédente en position non-limite (VC#) et de la

liquide initiale (OB#LI), mis en évidence chez les deux locuteurs expérimentaux, étaient reproduits ici en production isolée (V_1: 22.9%; C_3: 25.3%) comme en contexte de phrase (V_1: 20.5%; C_3: 18.6%). En outre, un léger allongement (7%) de l'obstruante en finale de mot (OB#LI), déjà observé chez le locuteur N.D., émergeait en production isolée, mais disparaissait en contexte de phrase. Comme précédemment, les séquences SOB montraient un léger allongement significatif (et non plus marginal) de l'obstruante à l'initiale de mot (de 6% en production isolée, et 5.2% en contexte de phrase). Par ailleurs, calquant le comportement de l'obstruante des groupes OBLI, le /s/ des séquences SOB montrait un léger allongement final en production isolée (6.7%), mais pas en contexte de phrase.

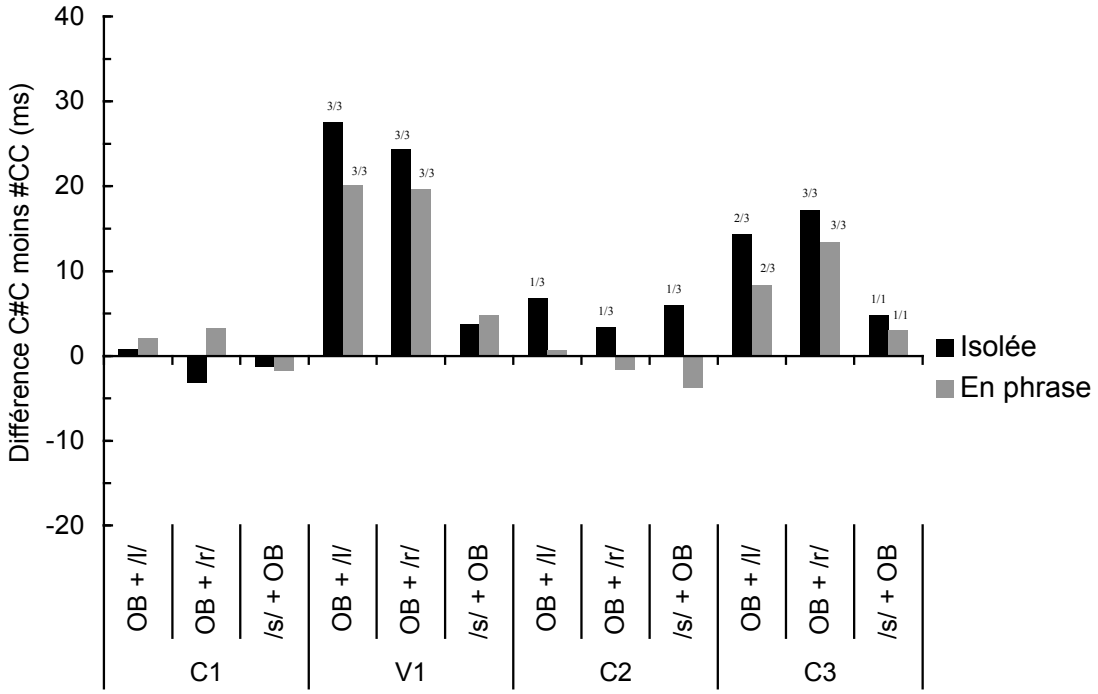

Figure 8. Différences de durée (ms) entre la condition #CC et la condition C#C, en fonction du contexte de réalisation et du groupe consonantique. Les différences significatives sont marquées d'un indice allant de 1 à 3 selon que le niveau de signification est inférieur à .05, .01, ou .001. L'indice de gauche est relatif à l'analyse par participant, celui de droite, à celle par item.

Par souci de clarté, les effets simples du groupe consonantique, effets purement liés aux durées phonémiques intrinsèques, et les interactions de ces effets avec ceux du contexte de réalisation ne seront pas rapportés. La Figure 8 présente la différence de durée entre la

condition C#C et la condition #CC, en fonction du contexte de réalisation et du groupe consonantique.

Consonne initiale. C_1 ne montrait aucun effet de la position de la frontière ou interaction de ce facteur avec le groupe consonantique ou avec le contexte de réalisation, et l'interaction double n'était pas significative (tous les *Fs* < 1 ou proches de 1, sauf pour l'interaction double par participant: $F_1(2,14) = 2.48$, $MSE = 10.1$, $p > .1$). Autrement dit, au bas de la hiérarchie, la condition C#C avait le même taux d'articulation que la condition #CC. En revanche, l'effet du contexte de réalisation était hautement significatif ($F_1(1,7) = 138.00$, $MSE = 206.5$, $p < .001$; $F_2(1,45) = 102.6$, $MSE = 555.6$, $p < .001$). En moyenne, lors du passage de production isolée à contexte de phrase, le raccourcissement subi par C_1 était de 36.1%.

Voyelle précédente. Sur V_1, l'effet de la position de la frontière ($F_1(1,7) = 47.37$, $MSE = 141.0$, $p < .001$; $F_2(1,45) = 37.49$, $MSE = 356.4$, $p < .001$) tout comme l'interaction de ce facteur avec le groupe consonantique étaient significatifs ($F_1(2,14) = 25.64$, $MSE = 36.6$, $p < .001$; $F_2(2,45) = 5.27$, $MSE = 356.4$, $p < .01$). Comme en témoignent des comparaisons planifiées, l'effet de la position de la frontière était significatif dans les groupes OBLI ($F_1(1,7) = 39.32$, $MSE = 115.2$, $p < .001$; $F_2(1,45) = 25.42$, $MSE = 356.4$, $p < .001$, pour la liquide /l/; $F_1(1,7) = 74.18$, $MSE = 52.4$, $p < .001$; $F_2(1,45) = 21.81$, $MSE = 356.4$, $p < .001$, pour la liquide /ʀ/), mais pas dans les groupes SOB ($F_1(1,7) = 3.05$, $MSE = 46.7$, $p > .1$; $F_2 < 1$). Les effets de la position de la frontière dans les deux conditions OBLI ne différaient pas statistiquement l'un de l'autre (les deux *Fs* < 1). En revanche, ils étaient tous deux significativement plus importants que dans la condition SOB ($F_1(1,7) = 33.69$, $MSE = 45.5$, $p < .001$; $F_2(1,45) = 8.61$, $MSE = 356.4$, $p < .01$, pour /l/ vs. SOB; $F_1(1,7) = 181.22$, $MSE = 7.0$, $p < .001$; $F_2(1,45) = 7.13$, $MSE = 356.4$, $p < .02$, pour /ʀ/ vs. SOB). L'effet du contexte de réalisation était significatif, V_1 étant 16.2% plus courte en contexte de phrase qu'en production isolée ($F_1(1,7) = 31.26$, $MSE = 307.4$, $p < .001$; $F_2(1,45) = 203.10$, $MSE = 94.6$, $p < .001$) et interagissait marginalement avec l'effet de la position de la frontière ($F_1(1,7) = 4.83$, $MSE = 16.6$, $p = .064$; $F_2(1,45) = 3.40$, $MSE = 47.3$, $p = .072$). Bien que moins marqué en contexte de phrase qu'en production isolée, l'effet de la position de la frontière était significatif dans les deux cas ($F_1(1,7) = 51.18$, $MSE = 80.4$, $p < .001$; $F_2(1,45) = 31.79$, $MSE =$

258.8, $p < .001$; $F_1(1,7) = 34.25$, $MSE = 77.3$, $p < .001$; $F_2(1,45) = 36.55$, $MSE = 144.8$, $p < .001$, respectivement). En revanche, l'interaction de second ordre n'était pas significative ($F_1(2,14) = 1.63$, $MSE = 23.1$, $p > .2$; $F_2(2,45) = 1.59$, $MSE = 47.3$, $p > .2$).

Première consonne du groupe. Sur C_1, ni l'effet de la position de la frontière ($F_1(1,7) = 1.28$, $MSE = 66.2$, $p < .3$; $F_2(1,45) = 2.57$, $MSE = 65.7$, $p < .15$, ni l'interaction groupe consonantique x position de la frontière n'étaient significatifs ($Fs < 1$ ou proches de 1). Toutefois, la présence d'un effet de la position de frontière dépendait du contexte de réalisation ($F_1(1,7) = 32.86$, $MSE = 8.7$, $p < .001$; $F_2(1,45) = 20.48$, $MSE = 27.8$, $p < .001$). Comme l'indiquent des comparaisons planifiées, l'effet de la position de la frontière était présent en production isolée ($F_1(1,7) = 8.64$, $MSE = 39.3$, $p < .03$; $F_2(1,45) = 12.52$, $MSE = 54.2$, $p < .001$), mais s'évanouissait complètement en contexte de phrase ($F_1 < 1$; $F_2(1,45) = 1.50$, $MSE = 39.2$, $p > .2$). L'effet du contexte de réalisation était significatif ($F_1(1,7) = 160.99$, $MSE = 36.1$, $p < .001$; $F_2(1,45) = 169.04$, $MSE = 68.8$, $p < .001$). Lors du passage au contexte phrastique, le raccourcissement subi par C_2 était de 19.7%. Enfin, l'interaction de second ordre n'était pas significative ($F_1(2,14) = 1.46$, $MSE = 8.1$, $p > .25$; $F_2 < 1$).

Seconde consonne du groupe. C_3 subissait des variations analogues à celles dont V_1 était l'objet. L'effet de la position de la frontière était hautement significatif ($F_1(1,7) = 58.40$, $MSE = 42.6$, $p < .001$; $F_2(1,45) = 85.11$, $MSE = 58.5$, $p < .001$), et interagissait avec l'effet du groupe consonantique ($F_1(2,14) = 11.54$, $MSE = 23.3$, $p < .002$; $F_2(2,45) = 9.2$, $MSE = 58.5$, $p < .001$). Bien que l'effet de la position de la frontière était significatif dans les trois conditions consonantiques ($F_1(1,7) = 26.23$, $MSE = 39.2$, $p < .002$; $F_2(1,45) = 35.21$, $MSE = 58.5$, $p < .001$, pour la liquide /l/; $F_1(1,7) = 54.29$, $MSE = 34.5$, $p < .001$; $F_2(1,45) = 64.10$, $MSE = 58.5$, $p < .001$, pour la liquide /ʀ/; $F_1(1,7) = 7.92$, $MSE = 15.4$, $p < .03$; $F_2(1,45) = 4.16$, $MSE = 58.5$, $p < .05$, pour les groupes SOB), il était nettement moins important dans la condition SOB ($F_1(1,7) = 7.80$, $MSE = 28.4$, $p < .03$; $F_2(1,45) = 7.58$, $MSE = 58.5$, $p < .01$, pour /l/ vs. SOB; $F_1(1,7) = 16.76$, $MSE = 31.1$, $p < .005$; $F_2(1,45) = 17.80$, $MSE = 58.5$, $p < .001$, pour /ʀ/ vs. SOB). En outre, l'effet de la position de la frontière était plus important pour /ʀ/ que /l/, mais uniquement par participant ($F_1(1,7) = 6.10$, $MSE = 10.3$, $p < .05$; $F_2(1,45) = 2.15$, $MSE = 58.5$, $p > .1$). L'effet du contexte de réalisation était significatif; en moyenne, C_3 subissait un

raccourcissement de 12.2% ($F_1(1,7) = 40.92$, $MSE = 43.2$, $p < .001$; $F_2(1,45) = 104.47$, $MSE = 33.8$, $p < .001$). En outre, le contexte de réalisation interagissait avec la position de la frontière ($F_1(1,7) = 12.19$, $MSE = 7.1$, $p < .02$; $F_2(1,45) = 13.77$, $MSE = 12.7$, $p < .001$). Bien que moins marqué en contexte de phrase qu'en production isolée, l'effet de la position de la frontière était significatif dans les deux cas ($F_1(1,7) = 62.57$, $MSE = 28.0$, $p < .001$; $F_2(1,45) = 90.34$, $MSE = 38.8$, $p < .001$; $F_1(1,7) = 37.83$, $MSE = 21.7$, $p < .001$; $F_2(1,45) = 50.89$, $MSE = 32.3$, $p < .001$). Enfin, l'interaction de second ordre n'était pas significative ($F_1 < 1$; $F_2(2,45) = 1.43$, $MSE = 12.7$, $p > .2$).

Au vu de ces résultats, on peut donc affirmer que le patron de variations des durées segmentales liées à l'assignation lexicale d'une consonne pivotale (i.e., C#C vs. #CC) dans les groupes consonantiques du français dépend de la nature du groupe consonantique. Pour l'essentiel, les variations de durée mises en évidence chez le locuteur C.H. se vérifient ici de manière hautement robuste et significative à travers d'autres locuteurs. Les groupes OBLI confirment les allongements substantiels de la voyelle précédente en position non-limite (VC#C), et de la liquide initiale (OB#LI), en production isolée comme en contexte de phrase. Ils suggèrent par ailleurs que le faible allongement de l'obstruante en finale de mot (OB#LI), émergeant en production isolée, et déjà observé chez le locuteur N.D., ne peut avoir de fonction de segmentation majeure, puisqu'il s'évanouit en contexte de phrase. Contrairement aux variations de la voyelle précédente et de la liquide, l'allongement de l'obstruante finale ne résiste donc pas au taux d'articulation plus élevé qu'implique l'enchâssement dans un contexte plus large (i.e., 5.3 syll/s vs. 3.7 syll/s en production isolée). Dans les groupes SOB, le segment /s/ est également touché par ce phénomène d'allongement final (/s/#OB). Ce dernier est donc indépendant de la nature du groupe consonantique. La seule autre variation est un léger allongement de l'obstruante initiale (/s/#OB), marginal chez le locuteur CH. et quatre fois plus faible que celui de la liquide OBLI. Toutefois, les Expériences 4 et 5 indiquent que cet allongement n'a aucune fonction perceptive dans la segmentation.

La mise en évidence de différences sur les durées segmentales en fonction de l'assignation lexicale de la consonne pivotale que ce soit sur les groupes OBLI ou les groupes SOB va radicalement à l'encontre de l'idée de resyllabation complète dans le cas de l'enchaînement (cf.

la discussion générale du chapitre 4). Logiquement, si les patrons articulatoires utilisés pour produire les séquences C#C étaient les mêmes que ceux employés dans la production des séquences alternatives #CC (Levelt *et al.* 1999; Levelt & Wheeldon, 1994), les mesures des durées segmentales auraient dû ne révéler aucune différence systématique. Toutefois, les données obtenues jusqu'ici ne permettent pas de dire si la resyllabation a lieu de manière incomplète, ou si elle n'a tout simplement pas lieu. Le but de l'expérience suivante était d'examiner cette question.

Expérience 6b – Comparaison entre frontières OBLI et frontières OBOB produites en contexte de phrases

L'Expérience 6b était similaire à l'Expérience 6, si ce n'est qu'elle comparait l'enchaînement dans les groupes OBLI (e.g., "tante roublarde") à une situation appariée dans laquelle la première consonne médiane ne pouvait être resyllabifiée avec la consonne suivante, puisqu'elles étaient toutes deux des consonnes obstruantes (e.g., "tante coudée"), et donc nécessairement hétérosyllabiques (i.e., "C.C"; Dell, 1995). Seule la réalisation en contexte de phrase était examinée. Si la condition OBLI diverge de la condition OBstruante-OBstruante (OBOB) en termes des durées segmentales, nous aurons un argument pour dire que la resyllabation a lieu partiellement; en revanche, l'absence de différence sera davantage compatible avec l'idée que la resyllabation est avant tout un *construct* théorique. Ici, comme l'identité de la seconde consonne du groupe médian variait d'une condition à l'autre, hormis la première consonne, seule la voyelle précédant la frontière (VC#C) offrait une mesure non-contaminée et donc interprétable.

Méthode

Participants. Seules deux locutrices ont été testées. Elles ne souffraient d'aucun trouble de l'audition ou du langage, et ne connaissaient pas les tenants ou les aboutissants de l'expérience. Aucune d'elles n'avait participé à l'expérience précédente.

Matériel et procédure. Le matériel —repris à l'Appendice 5.2— incluait les 32 phrases contenant les énoncés C#C à groupe OBLI utilisées dans l'Expérience 6 (e.g., "*Aurélie est une*

tante roublarde du Brabant"), et 32 phrases dérivées des premières en remplaçant l'adjectif (e.g., "*Aurélie est une tante coudée du Brabant*"). Celui-ci débutait par une obstruante de même voisement que l'obstruante à la finale du nom (e.g., "tant#cou"); avait le même nombre de syllabes et la même première voyelle que l'adjectif dans la condition OBLI (e.g., "roublarde" vs. "coudée"); et pour éviter autant que faire se peut les différences liées à la consonne suivante qui variait immanquablement, celle-ci avait le même trait de voisement dans les deux conditions (e.g., "roublarde" vs. "coudée"). Pour le reste, matériel et procédure étaient similaires à l'Expérience 6.

Résultats et discussion

Les durées segmentales moyennes en fonction de la nature du groupe consonantique sont présentées au Tableau 9 séparément pour chaque locutrice.

Tableau 9. Durées segmentales moyennes (ms) en fonction de la locutrice et de la nature du groupe consonantique. Les écarts-types figurent entre parenthèses.

Locutrice	$C_2\#C_3$	C_1	V_1	C_2	C_3	V_2
A.D.	OBLI	80 (30)	115 (36)	46 (17)	68 (24)	93 (35)
	OBOB	80 (32)	116 (31)	57 (19)	69 (22)	103 (36)
C.D.	OBLI	80 (27)	112 (35)	54 (24)	77 (22)	102 (35)
	OBOB	78 (25)	114 (45)	63 (31)	77 (22)	98 (35)

Note. Une cellule représente 32 données. L'identité de la consonne pivotale (C_3) variant d'une condition à l'autre, la durée de ce segment tout comme celles de la consonne précédente (C_2) et de la voyelle subséquente (V_2) ne sont rapportées qu'à titre indicatif.

Comme on peut le voir, si des variations difficilement interprétables affectaient la première consonne médiane et la seconde voyelle, la voyelle précédant la frontière lexicale (V_1), tout comme le premier segment du nom (C_1), ne montrait pratiquement aucune différence selon la nature du groupe consonantique, quelle que soit la locutrice examinée. Comme en témoignent des ANOVAs par item menées séparément pour chaque segment, le groupe consonantique et

la locutrice n'avaient aucun effet et n'interagissaient pas l'un avec l'autre, que ce soit sur C_1 ou sur V_1 (tous les $Fs < 1$).

Bien qu'il s'agisse d'un effet "nul", l'absence de différence sur la première voyelle entre les conditions OB#LI et OB#OB contraste fortement avec les différences observées précédemment sur cette même voyelle entre les conditions OB#LI et #OBLI dans les Expériences 1, 4 et 6. Pris dans leur ensemble, ces résultats sont donc davantage compatibles avec l'idée qu'il n'y a pas de resyllabation, même partielle, dans le cas de l'enchaînement en français.

Discussion générale

Ce chapitre avait pour but principal d'évaluer la nature générale des indices phonétiques de frontière lexicale mis en évidence au chapitre 4, en comparant les groupes OBLI à l'autre classe de groupes consonantiques potentiellement concernée par la question de la resyllabation, à savoir la classe SOB. Dans le présent chapitre, tout indique que la présence de tels indices et leur exploitation effective dans la segmentation lexicale sont conditionnées par la nature même du groupe consonantique. Dans le cas des groupes OBLI, les mesures des durées segmentales confirment, sur un total de neuf locuteurs naïfs, les variations initialement mises en évidence chez un locuteur non-naïf au chapitre précédent. Lorsque la frontière lexicale est située entre les deux consonnes du groupe médian, plutôt que devant l'obstruante, la voyelle précédente (VC#C) et la liquide (OB#LI) sont plus longues d'environ 20 à 25%, dépendant du taux d'articulation; à débit lent, l'obstruante est par ailleurs légèrement (7%) plus longue en finale de mot (OB#LI) qu'à l'initiale (#OBLI). En revanche, dans le cas des groupes SOB, ces mesures ne révèlent aucune variation équivalente de la voyelle ou de l'obstruante, mais un léger allongement du /s/ final (de 7%; /s/#OB) uniquement à débit lent, et de l'obstruante à l'initiale de mot (de 6%; /s/#OB).

En parallèle, les expériences de segmentation lexicale confirment dans le cas des groupes OBLI, l'exploitation de variations phonétiques —notamment sur les durées segmentales (cf. les corrélations)— liées à l'assignation lexicale de l'obstruante. Mais plus important ici, elles

démontrent qu'en dépit de légères variations sur le segment /s/ et sur l'obstruante, du point de vue des mécanismes de segmentation, les alternatives à groupe SOB sont homophones au sens strict du terme. En *word spotting*, le non-alignement final entre mots et intentions lexicales du locuteur induit un coût perceptif dans le cas des groupes OBLI (e.g., "tant#rou" vs. "tan#trou"), mais pas dans celui des groupes SOB (e.g., "ras#tu" vs. "ra#stu"). En tâche d'appariement co-déterminé, les participants parviennent à retrouver la segmentation lexicale sous-jacente des séquences OBLI, mais pas celle des séquences SOB. En outre, une analyse *post hoc* des scores d'appariement montre que les auditeurs restent "sourds" aux variations de durée dans les séquences SOB, même lorsqu'elles touchent la voyelle précédente et sont d'une magnitude équivalente à celles observées sur les OBLI. En somme, l'allongement systématique de la voyelle précédente et de la seconde consonne n'est pas une propriété générale des groupes consonantiques; mais en outre, il n'y a *a priori* aucun signe de généralisation perceptive: la capacité des auditeurs à exploiter ces allongements lorsqu'ils interviennent, dépend essentiellement de leur degré de systématicité dans la classe consonantique concernée.

L'absence d'indices exploitables des frontières lexicales dans les groupes SOB est assez intrigante. Etant donné les contraintes articulatoires impliquées par la friction, le segment /s/, comme les autres fricatives, est connu pour n'admettre qu'une faible marge de variations qualitatives (spectrales) comparé aux autres consonnes (Byrd, 1996), et donc on s'attendrait justement à ce qu'une dimension comme la durée soit, dans son cas, particulièrement utile pour encoder l'information prosodique. C'est le cas en anglais (Anderson & Port 1994; Byrd, 1996) et en néerlandais (Waals, 1999), où le /s/ des groupes SOB est nettement (environ 20 à 25%) plus long en début de syllabe (ou de mot) qu'à la finale.

Par ailleurs, les présentes données contrastent fortement avec celles rapportées récemment par Shatzman et McQueen (2006) en néerlandais. A l'aide de l'enregistrement des mouvements oculaires face à des images d'objets dénommés auditivement, ces auteurs ont trouvé que les néerlandophones distinguent des alternatives du type "eens pot" vs. "een spot" essentiellement sur base de la durée du /s/. Bien que ces données soient potentiellement contaminées par le fait que les alternatives n'étaient pas appariées en termes de structure

grammaticale, elles suggèrent que le comportement idiosyncratique des groupes SOB mis en évidence dans le présent chapitre est spécifique au français, voire aux langues romanes (cf. Bertinetto, 2004), et donc pas universel.

Bien que nous ne sachions pas ce qui motive l'absence d'indices de segmentation phonétique dans la classe SOB, nos résultats permettent, d'une certaine manière, d'expliquer l'extrême difficulté qu'ont les auditeurs des langues romanes à décider de l'assignation syllabique du phonème /s/ lorsqu'il est suivi d'une consonne (Bertinetto, 2004). En outre, ils convergent assez bien avec les descriptions phonologiques et phonétiques actuelles du français, qui donnent au /s/ un statut spécial. Rialland (1994) notamment, sur base de l'examen de processus tels que la formation des semi-voyelles, la gémination, et l'élision du schwa, indique que le phonème /s/ pourrait être indépendant de la structure syllabique, c'est-à-dire "extrasyllabique" (cf. Kaye, Lowenstamm, & Vergnaud, 1990). On peut donc penser que confrontée à un groupe SOB, l'heuristique de segmentation SOSH (Content *et al.*, 2001) ne privilégie aucun point de segmentation particulier: le /s/ comme l'obstruante seraient tous deux considérés comme des débuts de mots potentiels.

Enfin pour terminer, revenons à la question de la resyllabation. La mise en évidence de différences de durée systématiques entre la condition enchaînée (e.g., C#C) et la condition de consonnes initiales (#CC) peut signifier deux choses: soit que la resyllabation est incomplète, interprétation privilégiée par Fougeron *et al.* (2002), soit qu'elle n'a tout simplement pas lieu. L'Expérience 6b a tenté de répondre à cette question en comparant l'enchaînement à une situation dans laquelle la première consonne ne peut être resyllabifiée avec la consonne suivante puisqu'il s'agit d'obstruantes (e.g., "tant#rou" vs. "tant#cou"). En accord avec l'idée que la consonne pivotale n'appartient pas moins à la syllabe précédente dans le cas de l'enchaînement que dans la condition non-enchaînée, aucune différence de durée interprétable n'a été obtenue. Ces données suggèrent donc que l'allongement de la voyelle précédente en position limite et de la liquide initiale sont le reflet de l'assignation non-ambiguë de l'obstruante à la syllabe précédente.

Chapitre 6 Compétition lexicale et indices acoustico-phonétiques des frontières de mots

L'obtention dans les Expériences 1 et 4 d'un effet de l'alignement final entre mots et intentions lexicales (e.g., "tant#rou" vs. "tan#trou") de même magnitude (si pas plus important) que l'effet de l'alignement initial (e.g., ic#roche vs. i#croche) semble à première vue incompatible avec SOSH, l'heuristique de segmentation aux attaques syllabiques (Content *et al.* 2001; Dumay *et al.*, 2002). Puisque selon cette hypothèse, les débuts de syllabes constituent des points d'alignement privilégiés pour lancer les recherches lexicales, seul l'alignement de début entre mots et indices syllabiques devrait être déterminant pour l'activation lexicale. Cependant, comme nous l'avons dit dans la discussion de l'Expérience 1, l'incompatibilité entre SOSH et ces résultats n'est qu'apparente. Etant donné la composition lexicale des séquences porteuses (e.g., "tantrou"), il est fort probable que les cibles enchâssées à l'initiale (e.g., "tante") aient été systématiquement en compétition avec le mot CCV (e.g., "trou") pour l'attribution de la consonne pivotale (e.g., "t"). En effet, plusieurs études (Norris *et al.*, 1995; Vroomen & de Gelder, 1995) ont démontré que l'activation d'un mot enchâssé à l'initiale est d'autant plus faible qu'il y a de compétiteurs débutant sur le segment final de la cible (e.g., "melk" dans "melkaam" vs. "melkeum"), alors qu'aucun effet de ce genre n'est obtenu lorsque la cible et le set de candidats ne se chevauchent pas (e.g., "bel" in "belkaam" vs. "belkeum"; voir chapitre 2, section 1.2.1, pour plus de détails; cf. aussi la différence de latence entre les porteurs OBLI et SOB dans l'Expérience 4). Or, puisque ici, dans la condition non-alignée, le compétiteur chevauchant est particulièrement activé vu qu'il est aligné avec un début de syllabe, la consonne pivotale pourrait être davantage perçue comme lui appartenant, ce qui au final retarderait la détection du mot cible.

Le but du présent chapitre est de déterminer dans quelle mesure l'émergence d'effets d'alignement final en *word spotting* dépend de la compatibilité lexicale des portions d'input situées au-delà de la cible. Si, en accord avec SOSH, l'effet de non-alignement final (e.g., "tan#trou") résulte de l'influence inhibitrice accrue du compétiteur chevauchant, alors cet effet devrait disparaître lorsque la portion finale n'est pas un mot ou le début d'un mot plus long

(e.g., "tan#treum"). En revanche, si l'effet de non-alignement final est lié à la non-correspondance entre la représentation lexicale de la cible et les indices de fin de syllabe, que la portion finale soit un mot ou pas, ne devrait rien changer. L'Expérience 7 a évalué le rôle du compétiteur subséquent à l'aide des cibles CVC, puis l'Expérience 8 a cherché à confirmer les résultats à l'aide des cibles CV.

Expérience 7 – Effet d'alignement final et compétiteur chevauchant

Cette expérience examinait si l'effet d'alignement final entre mot et intentions lexicales du locuteur dépend, ou est accentué, par l'activation d'un compétiteur chevauchant la cible par la consonne pivotale. Des mots CVC ou CCVC étaient à détecter à l'initiale de porteurs C(C)VCCV ou C(C)VCCVC à segmentation lexicale ambiguë. Dans ces derniers, à la fois les indices acoustico-phonétiques de syllabation et le statut lexical de la portion finale étaient manipulés orthogonalement.

Méthode

Participants. Quarante-huit étudiants (âge moyen: 20 ans, limites: 18-27) de l'Université libre de Bruxelles ont participé à l'expérience pour l'obtention de crédits dans le cadre d'un cours d'introduction à la psycholinguistique. Tous étaient de langue maternelle française, et ont rapporté ne souffrir d'aucun trouble de l'audition ou du langage. Aucun d'eux n'avait participé aux expériences précédentes.

Matériel. Septante-deux mots français (noms ou adjectifs) dont 40 CVC (e.g., "tante") et 32 CCVC[24] (e.g., "flûte") ont été sélectionnés comme cibles (fréquence logarithmique moyenne: 3.30, limites: 1.08-5.01). La consonne finale de ces mots était une plosive ou une fricative et leur portion initiale CV était toujours un mot (e.g., "temps"; "flux"). Chaque cible apparaissait à l'initiale de deux types de séquences bisyllabiques à groupe consonantique médian OBLI (i.e., gl, kl, bʀ, dʀ, fʀ, pʀ, tʀ, et vʀ).

[24] A partir de maintenant, les cibles CVC et CCVC seront regroupées sous l'unique terme "CVC" pour rendre la description plus aisée.

Les effets combinés du non-alignement final et de la compétition lexicale étaient examinés à l'aide de porteurs de structure CVCCV (e.g., "tantrou"). Leur portion finale CCV était toujours un mot (i.e., "trou") et donc un compétiteur "chevauchant" de la cible initiale. Comme précédemment, deux versions auditives de chaque porteur ont été obtenues. Celles-ci étaient extraites d'énoncés de deux mots appariés, ne différant que par la position de la frontière lexicale (i.e., CVC#CV… comme *tante roublarde*, ou CV#CCV... comme *temps troublant*). Dans la condition alignée, la fin du mot cible coïncidait avec la frontière lexicale intentionnelle, tandis que le début du compétiteur CCV en était décalé d'un phonème (e.g., "tant#rou"). En revanche, dans la condition non-alignée, la fin du mot cible se trouvait décalée par rapport à la frontière lexicale intentionnelle, tandis que le début du compétiteur coïncidait avec cette dernière (e.g., "tan#trou").

Pour examiner l'effet du non-alignement en l'absence de compétition lexicale, les cibles étaient enchâssées dans des porteurs de structure CVCCVC (e.g., "tantreum"), tels que la liquide du groupe médian était la même que dans les porteurs à finale mot, et que les portions finales CCVC et CVC n'étaient pas des mots ou des débuts de mots. A nouveau, deux versions auditives de chaque porteur ont été extraites de paires minimales d'énoncés de trois syllabes, composés d'un nom monosyllabique et d'un pseudo-adjectif bisyllabique (i.e., CCVC#CVC… comme *tante reumlite*, ou CCV#CCVC… comme *temps treumlite*). Dans la condition alignée, la fin du mot cible coïncidait avec la frontière lexicale intentionnelle ("tant#reum"), tandis que dans la condition non-alignée, elle ne coïncidait pas avec cette dernière ("tan#treum"). Pour éviter que la consonne finale des porteurs CVCCVC ne soit annihilée par la coupure opérée lors de l'extraction des stimuli, seules des consonnes continues et/ou voisées ont été utilisées (i.e., /l/, /m/, /n/, /s/, /f/, /ʃ/, /z/, /ʒ/ et /g/). En outre, pour éviter que son identité ne soit modifiée par la consonne subséquente, laquelle était la même dans les deux conditions d'alignement, ces deux consonnes partageaient le même trait de voisement (e.g., /st/ dans "jambe reustive" vs. "gens breustifs").

Les remplisseurs consistaient en 40 séquences CVCCV(C) et 32 séquences CCVCCV(C), de même nature que les quadruplets expérimentaux, si ce n'est que leur portion initiale CVC ou CCVC n'était pas un mot (e.g., "chan#cleuf" extrait de *chant cleuftin*). Dans ces derniers,

les quatre conditions expérimentales (C#C/mot; #CC/mot; C#C/pseudomot; #CC/pseudomot) étaient représentées de manière équiprobable. Afin qu'il n'y ait aucune association entre l'identité du groupe médian et l'occurrence d'une cible, la distribution des groupes consonantiques au sein des remplisseurs était identique à celle des séquences porteuses. Par ailleurs, afin d'éviter tout effet d'amorçage, les mots CV enchâssés à l'initiale des porteurs, et la plupart des portions CVet CVC à l'initiale des remplisseurs, n'apparaissaient qu'une seule fois sur l'ensemble du matériel.

Les énoncés —repris aux Appendices 6.1 et 6.2— ont été produits en chambre sourde par une locutrice de langue maternelle française (L.D.), laquelle ignorait tout de l'expérience. Les conditions d'enregistrement étaient identiques à celles des Expériences 1 et 4. Les productions ont été transférées à une fréquence d'échantillonnage de 44.1 kHz et un taux de conversion D/A de 16 bits sur le disque dur d'un Macintosh Quadra-700. La procédure d'extraction des stimuli était identique à celle de l'Expérience 1, si ce n'est que pour les séquences à finale pseudomot, la coupe était opérée juste après la fin acoustique de leur consonne finale (e.g., "tantreum[lite").

Tableau 10. Durées segmentales moyennes (ms) en fonction de la position de la frontière lexicale (#) et du statut lexical de la portion finale CCV(C). Les écarts-types figurent entre parenthèses.

Conditions	C_1	C_{1b}	V_1	OB	LI	V_2	C_4
tant#rou	91 (60)	29 (41)	135 (36) ***	56 (19)	88 (15) ***	99 (30)	
tan#trou	93 (61)	28 (40)	120 (33)	55 (16)	73 (16)	102 (34)	
tant#reum	87 (57)	28 (40)	140 (36) ***	53 (15)	92 (15) ***	122 (14)	89 (19)
tan#treum	89 (55)	28 (39)	123 (31)	59 (15) **	77 (13)	119 (16)	87 (26)

Note. Une cellule (excepté C_{1b}) représente 72 données. C_{1b} désigne la seconde consonne (glide ou liquide) du groupe d'attaque dans les 32 cibles CCVC. Les durées des segments affectés par la procédure d'extraction des stimuli sont rapportées en gris. Les différences significatives en fonction de la position de la frontière lexicale sont notées "*" si $p < .05$, "**" si $p < .01$, et "***" si $p < .001$.

Afin de vérifier la présence des indices sur l'assignation de l'obstruante, les durées segmentales des porteurs ont été mesurées par l'auteur (voir Tableau 10), et chaque position

segmentale a fait l'objet d'une ANOVA avec le statut lexical de la portion finale CCV(C) (sauf pour C_4) et la position de la frontière lexicale (C#C vs. #CC) comme facteurs. En accord avec les résultats précédents (cf. Expériences 1 et 4), les porteurs montraient un allongement fiable et substantiel de la voyelle précédant la frontière en position non-limite (13.2%; $F(1,71) = 54.78$, $MSE = 349.1$, $p < .001$) et de la liquide initiale (20%; $F(1,71) = 145.44$, $MSE = 115.8$, $p < .001$), indépendant du statut lexical du second terme de l'énoncé (position de la frontière x statut lexical: $Fs < 1$). En outre, dans la condition pseudomot, l'obstruante était significativement plus longue (11.3%) à l'initiale du second terme qu'en finale du premier. L'interaction position de la frontière x statut lexical était hautement significative ($F(1,71) = 10.74$, $MSE = 81.5$, $p < .002$), et l'effet simple de la position de la frontière n'était significatif que dans la condition pseudomot ($F(1,71) = 11.65$, $MSE = 110.9$, $p < .002$; position de la frontière pour la condition mot: $F < 1$). L'assignation de l'obstruante était donc plus marquée dans la condition pseudomot que dans la condition mot.

Indépendamment de la position de la frontière lexicale, la première voyelle et la liquide étaient significativement plus longues (de 3.9 et 4.9%) dans les énoncés à finale pseudomot que dans ceux à finale mot ($F(1,71) = 6.03$, $MSE = 194.2$, $p < .02$; $F(1,71) = 8.76$, $MSE = 132.8$, $p < .005$, respectivement), ce qui suggère une articulation légèrement plus marquée dans les premiers. Enfin, notamment en raison de la coupure terminale du stimulus qui l'affectait dans la condition lexicale, la seconde voyelle était significativement plus courte (de 17%) dans les porteurs à finale mot qu'à finale pseudomot ($F(1,71) = 32.22$, $MSE = 956.6$, $p < .001$).

Quatre listes expérimentales comportant chacune 72 porteurs et 72 remplisseurs en ordre pseudo-aléatoire ont été composées. L'ordre des stimuli était tel que plus de trois porteurs et deux remplisseurs ne pouvaient apparaître en succession directe. Par ailleurs, si deux stimuli successifs requéraient la même décision ("oui, il y a un mot" vs. "non, il n'y en a pas"), ils appartenaient nécessairement à des conditions expérimentales différentes. La position des cibles et des remplisseurs était identique d'une liste à l'autre, le statut lexical de la portion finale CCV(C) et l'alignement étant contre-balancés à travers les listes. Une liste se scindait en deux parties en fonction de la structure phonologique des cibles (CVC vs. CCVC), et

chaque partie se décomposait à son tour en deux blocs équilibrés (2 x 40 essais puis 2 x 32 essais). Outre les listes expérimentales, deux blocs d'entraînement de 30 et 20 essais ont été constitués. Le premier, pour les cibles CVC, comprenait 20 porteurs; l'autre, pour les cibles CCVC en comprenait 12.

Procédure. Les participants ont été testés individuellement dans une pièce isolée. La présentation des stimuli, le timing de l'expérience et le recueil des données étaient contrôlés au moyen du logiciel PsyScope 1.2.4 (Cohen *et al.*, 1993) et d'un Macintosh-G3 Power PC. Les séquences de parole, stockées sur le disque dur de l'ordinateur, étaient présentées en monophonie binaurale à un volume d'écoute confortable, au moyen d'un amplificateur Kenwood KAF-3010R et d'écouteurs Beyerdynamic DT-770. La consigne était la même que dans la situation d'enchâssement initial de l'Expérience 1. Les participants ne pouvaient détecter que les mots se terminant par une et une seule consonne (e.g., "tante" mais pas "temps"). Les participants avaient 3 s pour répondre. Une fois le bouton pressé, un temps de réaction mesuré à partir du début de la séquence était enregistré. Chaque séquence était précédée d'un signal d'avertissement de 50 ms présenté 1 s plus tôt. Comme dans l'Expérience 3, l'intervalle entre les débuts de deux essais successifs dépendait directement du temps de réaction du participant, une période fixe de 2.5 s séparant la pression du bouton, du début de l'essai subséquent. En cas de non-réponse endéans les 3 s, le programme passait automatiquement à la période fixe de 2.5 s, puis lançait l'essai suivant. L'exactitude des productions orales des participants était contrôlée au vol par l'expérimentateur.

Douze participants ont été assignés à chaque liste. A l'intérieur d'une liste, la partie qui comportait les cibles CVC précédait systématiquement celle comportant les cibles CCVC. Toutefois, l'ordre de présentation des deux blocs constituant une partie était contre-balancé à travers les participants. Chaque partie était initiée par le bloc d'entraînement correspondant, celui dédié aux mots CVC étant présenté deux fois afin de familiariser les participants à la tâche. Entre les blocs expérimentaux, une pause d'une minute était proposée. L'expérience durait environ 40 minutes.

Résultats et discussion

Deux items (i.e., "bringue" et "fronde") omis par plus de 50% des participants ont été exclus des analyses. Les temps de réaction ont été ajustés en soustrayant la durée du CVC initial afin d'être calculés à partir de la fin de la cible. Les analyses chronométriques ont été réalisées sur les temps de réaction accompagnés d'une réponse orale correcte. La distribution de ces temps de réaction est rapportée pour chaque condition à l'Appendice 3 (Figure 11). Le taux d'erreurs moyen (i.e., omissions et réponses manuelles sans réponse orale ou accompagnées d'une réponse orale incorrecte) était de 12.1%. Le taux de fausses alarmes était de 3.4%. Le Tableau 11 présente les temps de réaction moyens (par participant) et les taux d'erreurs moyens pour l'Expérience 7.

Tableau 11. Temps de réaction moyens (TRs) mesurés à partir de la fin de la cible, et taux d'erreurs moyens (Ers) pour l'Expérience 7. Les écarts-types figurent entre parenthèses.

Mot		Type de Porteur	TRs (ms)	Ers (%)
Cible	Alignée	tant#rou	818 (201)	11.4 (9.8)
	Non-alignée	tan#trou	927 (217)	11.5 (8.9)
	Différence		109 *pi*	0.1
Pseudomot				
Cible	Alignée	tant#reum	895 (220)	13.6 (10.8)
	Non-alignée	tan#treum	934 (241)	11.7 (11.4)
	Différence		39	-1.9

Note. "*pi*" indique une différence statistiquement significative par participant et par item.

Les temps de réaction moyens par participant et item ont été soumis à des ANOVAs pour mesures répétées avec le statut lexical et l'alignement comme facteurs. L'effet du statut lexical était significatif ($F_1(1,47) = 7.65$, $MSE = 11111.1$, $p < .01$; $F_2(1,69) = 5.04$, $MSE = 21305.4$, $p < .03$). Les porteurs de la condition lexicale étant plus courts d'un phonème mais aussi plus courts au niveau des deux voyelles et de la liquide (cf. Tableau 10), les latences de détection étaient plus rapides pour les porteurs à finale mot que pour ceux à finale pseudomot (872 ms

vs. 915 ms). L'effet de l'alignement était hautement significatif ($F_1(1,47) = 25.19$, $MSE = 10297.0$, $p < .001$; $F_2(1,69) = 11.02$, $MSE = 37095.6$, $p < .002$). En moyenne, les cibles étaient détectées 73 ms plus rapidement lorsque leur fin était alignée avec la frontière lexicale intentionnelle que lorsqu'elle ne l'était pas (857 ms vs. 930 ms). Mais plus important, l'émergence de l'effet d'alignement final dépendait de la présence d'un compétiteur chevauchant, comme en témoigne l'interaction significative entre le statut lexical et l'alignement ($F_1(1,47) = 4.04$, $MSE = 14818.0$, $p = .05$; $F_2(1,69) = 6.02$, $MSE = 25096.4$, $p < .02$). Dans le cas des porteurs à finale mot, l'alignement avait un effet hautement significatif: les latences de détection étaient en moyenne 109 ms plus courtes dans la condition alignée que dans la condition non-alignée ($F_1(1,47) = 26.95$, $MSE = 10544.7$, $p < .001$; $F_2(1,69) = 19.44$, $MSE = 27173.5$, $p < .001$). En revanche, bien que les latences de détection étaient légèrement plus courtes (39 ms) dans la condition alignée que dans la condition non-alignée, l'effet de l'alignement sur les porteurs à finale pseudomot était loin du seuil de signification ($F_1(1,47) = 2.41$, $MSE = 14570.3$, $p > .1$; $F_2 < 1$). En ce qui concerne les erreurs, les taux étaient sensiblement les mêmes à travers les quatre conditions. Des ANOVAs similaires à celles effectuées sur les temps de réaction ne révélaient aucun effet ou interaction significative (tous les $Fs < 1$).

La mise en évidence d'un effet de l'alignement sur les porteurs à finale mot (e.g., tant#rou vs. "tan#trou") atteste de la pertinence perceptive des différences observées sur les durées segmentales (ou de quelque corrélat de ces différences), et reproduit les résultats des Expériences 1 et 4. Mais plus important, le fait que cet effet se réduise à une tendance non-significative en l'absence de compétiteur chevauchant (e.g., tant#reum vs. "tan#treum") argue fortement en faveur de l'idée que l'effet de non-alignement final est imputable à la compétition accrue entre la cible CVC et le mot CCV activé par l'attaque subséquente, comme le prédit SOSH. Notons que la tendance non-significative sur les porteurs à finale pseudomot pourrait aussi être le reflet de cette dynamique, toutefois très vite avortée en raison de l'incompatibilité lexicale de leur portion CCVC.[25]

[25] La disparition de l'effet d'alignement final en l'absence de compétiteur chevauchant démontre à elle seule que les effets d'alignement ne sont pas réductibles à des différences de temps d'input, argument gardé secret dans la discussion de l'Expérience 1.

Comme dans les expériences précédentes, nous avons examiné les corrélations entre la différence de durée sur la première voyelle, l'obstruante et la liquide, et l'effet de l'alignement final (i.e., C#C moins #CC) sur les latences de détection. Contrairement aux Expériences 1 et 4, dans lesquelles aucune des corrélations examinées n'était significative, ici, et peut-être est-ce lié au nombre d'items deux fois plus important, la liquide des porteurs à finale mot, pourtant plus longue dans la condition alignée que non-aligné, montrait une légère corrélation positive, autant inattendue qu'ininterprétable (r(70) = liquide: .27, $p < .05$). En revanche, les autres corrélations étaient toutes non-significatives (tous les rs(70) entre -.16 et .20, $ps \geq .1$).[26] Sur base uniquement de la situation d'enchâssement initial, il s'avérerait donc assez difficile de relier sur le plan statistique la performance de segmentation et les variations des durées segmentale (mais voir les Expériences 1 et 2).

L'Expérience 8 a poursuivi l'exploration de la dynamique de compétition lexicale dans ces séquences ambiguës à l'aide de la détection des mots CV (ou CCV) enchâssés à l'initiale. A partir de maintenant, les conditions alignée et non-alignée étaient donc renversées. Dans le cas qui nous occupe (e.g., "tan#trou" vs. "tant#rou"), la cible n'est jamais interrompue par la frontière lexicale, et les candidats lexicaux compatibles avec la portion finale CCV ne sont jamais en compétition avec la cible puisqu'ils ne la chevauchent pas. Cependant, comme on vient de le voir, dans la condition alignée, la portion finale CCV, lorsqu'elle est un mot (e.g., "trou"), inhibe très clairement la reconnaissance du mot CVC à l'initiale (e.g., "tante"). Or, si ce dernier est un compétiteur du mot CV (e.g., "temps"; cf. Vroomen & de Gelder, 1997a), le mot CCV final est plutôt un allié de la cible (cf. Frauenfelder & Peeters, 1990). Par conséquent, sur base de cette logique, on s'attend à obtenir la même interaction entre compatibilité lexicale et alignement que celle trouvée dans l'Expérience 7, avec un effet de non-alignement de forte magnitude lorsque la portion finale CCV est un mot.

[26] Il n'y avait par ailleurs aucune corrélation négative entre les différences de durée de cible (alignée moins non-alignée) et l'effet d'alignement final (mot: r(70) = -.10, $p > .4$ sur les TRs; r(70) = .08, $p > .6$ sur les Ers; pseudomot: r(70) = -.13, $p > .2$ sur les TRs; r(70) = .18, $p > .1$ sur les Ers).

Expérience 8 – Effet d'alignement final, compétiteur enchâssant, et compétiteur du compétiteur

Méthode

Participants. Quarante-huit étudiants (âge moyen: 21 ans, limites: 18-31) de l'Université libre de Bruxelles ont participé à l'expérience pour l'obtention de crédits dans le cadre d'un cours d'introduction à la psycholinguistique. Tous étaient de langue maternelle française, et ont rapporté ne souffrir d'aucun trouble de l'audition ou du langage. Aucun d'eux n'avait participé aux expériences précédentes.

Matériel. Le matériel expérimental était celui de l'Expérience 7. Toutefois, étant donné qu'ici, ce sont les mots CV ou CCV enchâssés à l'initiale qui faisaient office de cibles, d'autres remplisseurs —repris à l'Appendice 6.3— tels que leur portion initiale CV ou CCV ne constituait pas un mot, ont été créés (e.g., "dat#reug" extrait de *date reugmache*). En raison du nombre limité de mots CVC dont la portion initiale CV n'est pas un mot (i.e., 31 mots), la proportion de remplisseurs de structure CVCCV(C) n'était pas toujours de 10/18. Pour le reste, le matériel était similaire à l'Expérience 7.

Procédure. Les participants avaient pour consigne de détecter les mots enchâssés à l'initiale pourvu qu'ils ne terminaient point par une consonne. Le reste de la procédure était identique à l'Expérience 7.

Résultats

Six items (i.e., "baie", "cas", "clos", "flux", "grès" et "tes") omis par plus de 50% des participants ont été exclus des analyses. Les temps de réaction ont été ajustés en soustrayant la durée du CV initial afin d'être calculés à partir de la fin de la cible. Les analyses chronométriques ont été réalisées sur les temps de réaction accompagnés d'une réponse orale correcte. La distribution de ces temps de réaction est rapportée pour chaque condition à l'Appendice 3 (Figure 12). Le taux d'erreurs moyen (i.e., omissions et réponses manuelles sans réponse orale ou accompagnées d'une réponse orale incorrecte) était de 13.6%. Le taux de fausses alarmes était de 9.6%. Le Tableau 12 présente les temps de réaction moyens (par participant) et les taux d'erreurs moyens pour l'Expérience 8.

Tableau 12. Temps de réaction moyens (TRs) mesurés à partir de la fin de la cible, et taux d'erreurs moyens (Ers) pour l'Expérience 8. Les écarts-types figurent entre parenthèses.

Mot		Type de Porteur	TRs (ms)	Ers (%)
Cible	Alignée	tan#trou	1012 (208)	12.8 (10.7)
	Non-alignée	tant#rou	1037 (204)	12.9 (10.7)
	Différence		25 $^{p(i)}$	0.1
Pseudomot				
Cible	Alignée	tan#treum	1083 (251)	14.0 (10.8)
	Non-alignée	tant#reum	1161 (241)	14.8 (11.3)
	Différence		78 $^{p(i)}$	0.8

Note. L'effet de l'alignement sur les temps de réaction était statistiquement significatif par participant et frôlait le seuil de signification par item, d'où l'utilisation de l'indice "$p(i)$".

Les temps de réaction ont été analysés au moyen d'ANOVAs avec le statut lexical et l'alignement comme facteurs. L'effet du statut lexical était significatif ($F_1(1,47) = 20.70$, $MSE = 22140.9$, $p < .001$; $F_2(1,65) = 24.42$, $MSE = 22145.4$, $p < .001$). Les latences de détection étaient à nouveau plus courtes (de 98ms) pour les porteurs à finale mot que pour ceux à finale pseudomot (1024 ms vs. 1122 ms). L'effet de l'alignement (52 ms) était significatif par participant et au seuil de signification par item (1047 ms vs. 1099 ms; $F_1(1,47) = 7.91$, $MSE = 16230.8$, $p < .01$; $F_2(1,65) = 3.80$, $MSE = 25473.5$, $p = .055$). Toutefois, contrairement à ce qu'on attendait, l'interaction entre le statut lexical et l'alignement n'était pas significative ($F_1(1,47) = 2.01$, $MSE = 16412.8$, $p > .15$; $F_2(1,65) = 1.77$, $MSE = 30847.8$, $p > .15$). Descriptivement, l'effet de l'alignement était d'ailleurs moins marqué pour les porteurs à finale mot que pour ceux à finale pseudomot. Les taux d'erreurs étaient très similaires à travers les quatre conditions. Des ANOVAs analogues à celles effectuées sur les temps de réaction ne révélaient aucun effet ou interaction significative.

Analyses jointes des Expériences 7 et 8

Les données des Expériences 7 et 8 ont été analysées conjointement au moyen d'ANOVAs pour mesures répétées incluant la structure phonologique des cibles (CVC vs. CV), le statut lexical de la portion finale (mot vs. pseudomot) et l'alignement (alignée vs. non-alignée) comme facteurs. Dans les analyses par item, les cellules vides liées à l'exclusion d'un item dans l'une ou l'autre expérience étaient comblées par la donnée moyenne calculée en fonction du statut lexical pour l'expérience concernée. Outre les effets significatifs du statut lexical ($F_1(1,94) = 28.21$, $MSE = 16626.0$, $p < .001$; $F_2(1,71) = 28.05$, $MSE = 21479.7$, $p < .001$) et de l'alignement ($F_1(1,94) = 28.38$, $MSE = 13263.9$, $p < .001$; $F_2(1,71) = 13.23$, $MSE = 32565.4$, $p < .001$), l'analyse des temps de réaction confirmait, par l'interaction double structure phonologique x statut lexical x alignement, la différence nette entre les deux patrons de résultats ($F_1(1,94) = 5.82$, $MSE = 15615.4$, $p < .02$; $F_2(1,71) = 7.60$, $MSE = 24228.9$, $p < .01$). Par l'interaction structure phonologique x statut lexical ($F_1(1,94) = 4.47$, $MSE = 16626.0$, $p < .04$; $F_2(1,71) = 4.80$, $MSE = 19500.7$, $p < .04$), elle attestait du renversement complet —toute proportion gardée— de l'influence tantôt inhibitrice tantôt facilitatrice de la lexicalité de la portion finale, selon que la cible et le mot CCV se chevauchaient ou pas dans le temps. Enfin, par la présence d'un effet significatif de la structure phonologique ($F_1(1,94) = 20.04$, $MSE = 154560.9$, $p < .001$; $F_2(1,71) = 22.50$, $MSE = 190124.8$, $p < .001$) en l'absence de tout échange entre rapidité et précision (13.6% vs. 12.1%; $Fs <$ ou proches de 1), elle démontrait que les mots CV sont plus difficiles à détecter que les mots CVC. Pour le reste, les interactions simples impliquant l'alignement n'étaient pas significatives (tous les $Fs < 1$, sauf pour structure phonologique x alignement: $F_2(1,71) = 2.04$, $MSE = 27079.6$, $p > .15$)[27]. Des ANOVAs similaires effectuées sur les taux d'erreurs ne révélaient aucun effet ou interaction significative.

[27] A l'instar des autres expérience, l'Expérience 8 ne montrait aucune corrélation négative entre les différences de durée de cible et l'effet d'alignement final (mot: $r(66) = -.03$, $p > .8$ sur les TRs; $r(66) = .09$, $p > .4$ sur les Ers; pseudomot: $r(66) = .10$, $p > .4$ sur les TRs; $r(66) = .16$, $p > .15$ sur les Ers).

Discussion générale

De toute évidence, les résultats de l'Expérience 8 contrastent fortement avec ceux de l'Expérience 7 et ne collent pas du tout avec nos prédictions. Premièrement il n'y a pas d'interaction entre l'alignement et le statut lexical de la portion finale CCV(C). Contrairement à nos attentes, l'effet de l'alignement est présent aussi bien pour les porteurs à finale pseudomot que pour ceux à finale mot. Il est par conséquent difficilement explicable par la modulation du phénomène de compétition lexicale qui oppose le mot CVC enchâssant la cible à la portion finale CCV (cf. Expérience 7). Néanmoins, deux explications alternatives peuvent être proposées. L'une est que les variations acoustico-phonétiques mises en évidence, dont notamment la différence de durée sur la première voyelle, modulent directement l'activation des candidats lexicaux en fonction de la localisation de la frontière subséquente. Des indices de ce type ont été rapportés en ce qui concerne la distinction entre les mots mono- et multisyllabiques, tels que "chat" et "chagrin" (cf. Christophe, 1993; Davis *et al.*, 2002; Salverda *et al.*, 2004; cf. chapitre 2, section 2.1.2.2). Toutefois ici, il s'agirait d'éléments liés non pas à l'occurrence (i.e., présence vs. absence) d'une frontière lexicale entre la voyelle qui varie et celle qui suit, mais à la position phonémique exacte d'une frontière lexicale dont on sait qu'elle intervient quelque part entre ces deux voyelles. La présence d'un effet de l'alignement de 39 ms sur les porteurs à finale pseudomot dans l'Expérience 7, bien que non-significatif, pourrait aussi être due à ce mécanisme. Malheureusement l'absence, ici comme dans les expériences précédentes, de corrélation négative entre les différences de durée sur la première voyelle, l'obstruante et la liquide (C#C moins #CC), et l'effet de l'alignement final (V_1: $r(32) = .08$, $p > .5$; obstruante: $r(32) = .01$, $p > .9$; liquide: $r(32) = -.14$, $p > .2$, pour les porteurs à finale mot; V_1: $r(32) = .09$, $p > .4$; obstruante: $r(32) = -.09$, $p > .4$; liquide: $r(32) = -.11$, $p > .4$, pour les porteurs à finale pseudomot) ne permet pas de soutenir davantage cette hypothèse. Par ailleurs, cette dernière est incompatible avec les résultats de l'analyse corrélationnelle présentée au chapitre 1, qui suggère que les indices phonétiques liés à la position des frontières lexicales sont perçus uniquement comme des indices de début de mot (ou de syllabe).

L'autre explication fait appel à l'hypothèse PWC introduite par Norris *et al.* (1997; voir

aussi Cutler *et al.*, 2001, 2002; Norris *et al.*, 2001; cf. chapitre 2, section 3). D'après l'heuristique proposée, tout candidat lexical potentiellement présent dans le continuum serait désactivé, voire inhibé, si l'une des deux portions de parole qui s'étendent de ses limites gauche et droite à une frontière de mot hautement probable ne peut elle-même pas constituer un mot, par exemple, en raison de l'absence de voyelle (Cutler *et al.*, 2002). Dans les présentes conditions de non-alignement (e.g., "tant#reum"), la frontière lexicale la plus probable est située juste avant la liquide. Par conséquent, reconnaître le mot CV enchâssé à l'initiale ne pourrait donc pas être viable puisque cela équivaudrait à laisser l'obstruante sans interprétation lexicale. Autrement dit "temps" aurait son activation réduite aussi bien dans "tant#rou" que dans "tant#reum", parce que le segment "t" situé entre sa limite droite et la frontière lexicale qui a été perçue entre les consonnes du groupe médian n'a pas le format d'un mot. En somme dans le cas qui nous occupe ici, l'application de l'heuristique PWC impliquerait des connaissances sur les structures phono-lexicales possibles en français, combinées avec la détection des frontières lexicales hautement probables sur base de SOSH (Content *et al.*, 2001). Sous cet angle, les effets d'alignement obtenus dans l'Expérience 8, bien qu'inattendus, ne contrediraient donc pas l'hypothèse de segmentation aux attaques syllabiques.

Le second résultat intrigant dans l'Expérience 8 est la présence sensible d'une tendance à l'interaction entre le statut lexical et l'alignement, avec un effet de l'alignement en moyenne plus important pour les porteurs à finale pseudomot. Il y a ici aussi au moins deux manières d'expliquer ce phénomène. Comme le suggère l'allongement de l'obstruante en position initiale dans les porteurs à finale pseudomot (cf. Tableau 10), ces derniers pourraient présenter des indices de frontière lexicale plus puissants que les porteurs à finale mot.

Alternativement, la réduction de l'effet de non-alignement dans la condition mot pourrait refléter une interaction beaucoup plus complexe entre les différentes sources d'information impliquées dans PWC que ne l'ont jusqu'à présent envisagé Norris et coll. D'après ces auteurs, l'occurrence d'une portion de parole non-viable du point de vue phono-lexical conduit à la désactivation des candidats lexicaux adjacents, mais ne modifie en aucun cas la décision du système quant au point de segmentation le plus probable. Etant donné la tendance à

l'interaction entre le statut lexical et l'alignement telle qu'observée dans l'Expérience 8, il nous semble au contraire que le système de traitement révise ses choix de segmentation lorsque le rattachement de la portion non-interprétée avec ce qui est situé au-delà du point de segmentation initialement perçu est viable sur le plan lexical. Autrement dit, le non-alignement de "temps" dans "tante" est plus facile à surmonter dans "tant#rou" que dans "tant#reum" parce que "trou" est un mot, mais "treum" ne l'est pas. Cependant, avec les données dont nous disposons, nous ne pouvons dire si ce phénomène est dépendant de l'analyse lexicale intentionnelle ou bien s'il est automatique.

Chapitre 7 Conclusion générale

Cette recherche a examiné du côté de la production, et de celui de la perception de la parole, la présence d'indices acoustico-phonétiques sur la position des frontières lexicales dans les groupes consonantiques du français. Des paires de séquences à segmentation lexicale ambiguë (e.g., "tant#rou" vs. "tan#trou") étaient extraites d'énoncés à segmentation non-ambiguë, appariés notamment en structure grammaticale et en patron accentuel (e.g., "tante roublarde" vs. "temps troublant"). De par leur nature non-ambiguë, ces énoncés permettent d'étudier l'influence de corrélats phonétiques des frontières de mots *de facto* non-spécifiques aux paires minimales. Pour la classe des groupes consonantiques OBLI, nous avons mis en évidence un allongement substantiel, et systématique à travers les locuteurs, de la voyelle précédente en position non-limite (VC#C) et de la liquide initiale (OB#LI) (cf. Expériences 1, 4, 6 et 7). Ces phénomènes émergent non seulement en production isolée, mais aussi en contexte de phrase (cf. Expérience 6), et ils sont indépendants du statut lexical du second terme (e.g., "tant#roublarde" vs. "tant#reumlite"; cf. Expérience 7). Ce dernier élément indique que, du moins dans le cas de la liquide, l'allongement observé n'est donc pas une propriété des représentations lexicales, mais un phénomène purement prosodique, et dès lors, post-lexical. Par ailleurs, nous avons trouvé un allongement faible et optionnel de l'obstruante finale (OB#LI), uniquement en production isolée (cf. Expériences 1 et 6).

Sur le front de la perception des groupes OBLI, les quatre tâches utilisées (i.e., *word spotting*, catégorisation, répétition partielle et appariement co-déterminé) confirment la présence de co-variations phonétiques associées à la position de la frontière lexicale. Mais plus important, elles démontrent que l'auditeur exploite ces indices pour segmenter les énoncés en leurs constituants lexicaux, ou encore pour déterminer la syllabation de l'entrée acoustique (cf. Expériences 1-5, 7 et 8). Par exemple, dans la tâche de *word spotting*, nous avons mis en évidence l'existence de coûts de traitement liés au non-alignement initial, ou final, entre mots et intentions lexicales ("i#croche", cf. Expérience 1; "tan#trou", cf. Expériences 1, 4 et 7; "tant#rou", cf. Expérience 8). Nous avons montré aussi que l'émergence

de l'effet de non-alignement final est conditionnée par la présence d'un compétiteur chevauchant (cf. Expérience 7).

En outre, des corrélations entre la performance et les variations des durées segmentales indiquent que l'auditeur exploite directement la durée de la voyelle précédente et celle de la liquide pour localiser les frontières de mots. Les choix de segmentation lexicale (en tâche de catégorisation) s'avèrent sensibles aux deux paramètres. En revanche, la segmentation *online*, examinée ici à travers la performance de *word spotting*, semble utiliser de manière prépondérante la durée de la liquide, c'est-à-dire la durée du segment initial (cf. Figure 3). Enfin, une analyse corrélationnelle entre les tâches suggère que les variations phonétiques observées ne sont pas perçues comme des indices sur la position des frontières entre les mots, ou les syllabes, mais uniquement comme des indices sur la localisation des débuts de mots, ou de syllabes (cf. chapitre 4).

La mise en évidence de différences sur les durées segmentales (tout comme sur la structure formantique; Content *et al.*, 2004; Fougeron *et al.*, 2002) entre les conditions C#(C)V et #C(C)V suggère que l'enchaînement en français ne conduit pas à la resyllabation de la chaîne phonémique, ou du moins pas complètement. L'obtention d'effets perceptifs liés à l'assignation lexicale de l'obstruante, dans les quatre tâches examinées, confirme cette interprétation. En outre, l'Expérience 6b suggère que les conditions OB#LI et OB#OB (e.g., "tante roublarde" vs. "tante coudée") ont la même structure syllabique (C.C). Par conséquent, l'ensemble des données est davantage compatible avec l'idée que, dans le cas de l'enchaînement, la resyllabation n'a pas lieu, même partiellement. Contrairement aux hypothèses avancées par Levelt et coll. (Levelt, 1989; Levelt *et al.*, 1999; Levelt & Wheeldon, 1994), l'encodage phonologique garde donc des traces de la position des frontières entre les mots qui composent un même empan articulatoire.

En revanche, sur les groupes consonantiques SOB, nous n'avons trouvé aucune différence de durée segmentale équivalente en termes de magnitude moyenne, et de systématicité, à celles observées sur les groupes OBLI. Tout au plus les groupes SOB montrent-ils un léger allongement de l'obstruante initiale (/s/#OB), que ce soit en production isolée ou en contexte de phrase, ainsi qu'un léger allongement du /s/ final (/s/#OB), uniquement en production

isolée (cf. Expériences 4 et 6). La force et la systématicité des variations phonétiques liées à la position des frontières de mots dans les groupes consonantiques du français dépend donc de la nature du groupe consonantique.

Les groupes SOB contrastent fortement avec les groupes OBLI, aussi sur le front de la perception. Les deux tâches de segmentation utilisées (i.e., *word spotting* et appariement co-déterminé) indiquent que les groupes SOB ne présentent aucun indices exploitables sur l'assignation lexicale du /s/; ceci est vrai, en particulier, même lorsque l'on restreint l'analyse de la performance aux seules séquences qui montrent des variations équivalentes à celles observées sur les groupes OBLI (cf. Expériences 4 et 5). En somme, la capacité des auditeurs à exploiter effectivement les différences phonétiques liées à la position des frontières de mots est, elle-même, conditionnée par le degré de systématicité de ces phénomènes dans la classe consonantique concernée.

Dans l'ensemble, les données obtenues, notamment sur les groupes OBLI, sont compatibles avec l'idée que les attaques de syllabes constituent des points d'alignement privilégiés pour les recherches lexicales, comme le propose SOSH (Content *et al.*, 2001; Dumay *et al.*, 2002). Cependant, sur base des expériences rapportées ici, il est difficile de dire —et le lecteur l'aura remarqué— si les indices acoustico-phonétiques mis en évidence sont des indices de début de syllabe ou des indices de début de mot. Le cas le plus probable est qu'il s'agit d'une combinaison d'indices des deux types. Comme nous l'avons vu dans l'introduction (chapitre 2, section 2.1), les allophones de début de mot ne diffèrent pas des allophones de début de syllabe, que ce soit sur le plan articulatoire (Fougeron, 1998, pour le français; Keating *et al.*, 1999; Krakow, 1989, pour l'anglais) ou sur le plan acoustique (Fougeron *et al.*, 2002; Schaegis *et al.*, 2005, pour le français). L'allongement du segment à l'initiale de mot est donc à considérer avant tout comme un indice de début de syllabe, pas comme un indice de début de mot. En revanche, les variations sur la voyelle précédente, plus longue en syllabe finale qu'en syllabe initiale ou médiane (e.g., "c<u>a</u>r fourgué" vs. "c<u>a</u>rrefour gai"), reflètent avant tout la nature lexicale de la frontière avec la syllabe subséquente (Christophe, 1993). Il se pourrait donc que l'allongement de la voyelle finale en position non-limite (<u>V</u>C#) soit un indice de syllabation spécifique des frontières de mots.

Sous cet angle, l'application de l'heuristique SOSH pourrait donc être strictement dépendante de la puissance des indices phonétiques présents sur la voyelle précédente. Lorsque ces indices reflètent suffisamment le fait que la frontière subséquente n'est pas une frontière lexicale, nul besoin de recourir à SOSH —le but d'une heuristique de segmentation n'est pas d'engendrer gratuitement de l'activité lexicale. Alternativement, la détection des attaques syllabiques, que celles-ci correspondent ou pas à des débuts de mots, pourrait être systématique. Ici, c'est l'activation lexicale induite sur l'attaque repérée qui pourrait être modulée par la puissance des indices quant à la probabilité qu'il s'agisse d'une frontière de mot. Par exemple, lorsque ces indices informent suffisamment qu'il ne s'agit pas d'une frontière lexicale, l'activation lexicale engendrée par SOSH serait conséquemment fortement réduite. Contrairement à la première, cette dernière hypothèse prédit donc une augmentation graduelle de la différence entre les conditions alignée et non-alignée avec l'augmentation de la puissance des indices situés sur la première voyelle.

Enfin, la présente recherche s'inscrit directement dans ce courant d'études soulignant l'importance des détails infra-phonémiques dans le traitement lexical. Dans l'introduction, nous avons vu que certains de ces détails guident l'appariement au lexique en fournissant des indices coarticulatoires sur l'identité des segments (e.g., Streeter & Nigro, 1979; Dahan *et al.*, 2001). Ici, nous montrons que la représentation d'entrée contient également, dans certains cas, de subtils indices sur la position des frontières de mots. La fonction de ces indices pourrait être d'aider l'auditeur à déterminer sur quels endroits du signal allouer un maximum de ressources de traitement afin d'extraire les mots du continuum.

Bibliographie

Allopenna, P.D., Magnuson, J.S., & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, **38,** 419-439.

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

Altmann, G.T.M. (1999). Thematic role assignment in context. *Journal of Memory and Language*, **41,** 124–145.

Altmann, G.T.M., & Carter, D. (1989). Lexical stress and lexical discriminability: stressed syllables are more informative, but why? *Computer Speech and Language*, **3,** 265-275.

Altmann, G.T.M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, **73,** 247-264.

Anderson, S. (1982). The analysis of French schwa: or, how to get something for nothing. *Language*, **58,** 534-573.

Anderson, S., & Port, R. (1994). Evidence for syllable structure, stress, and juncture from segmental durations. *Journal of Phonetics*, **22,** 283-315.

Andruski, J.E., Blumstein, S.E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, **52,** 163-187.

Autesserre, D., & Chafcouloff, M. (1998). L'extraction et la mise en relation des indices acoustiques de /ʀ/ en français: Première étape d'une analyse phonétique intégrée. *Travaux de l'Institut Phonétique d'Aix*, **18,** 13-35.

Baayen, R.H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, **36,** 94-117.

Balota, D.A., & Chumbley, J.I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, **10,** 340-357.

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.

Banel, M.-H., Frauenfelder, U.H., & Perruchet, P. (1998). Contribution des indices de métriques à l'apprentissage d'un langage artificiel. *Actes des XXIIe Journées d'Etude sur la Parole* (pp. 29-32). Martigny, Suisse.

Bard, E.G., Shillcock, R.C., & Altmann, G.T.M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception & Psychophysics*, **44,** 395-408.

Barry, W.J. (1981). Internal juncture and speech communication. *Arbeitsberichte Institut für Phonetik Universität Kiel*, **16,** 229-287.

Barry, W.J. (1984). Perception of juncture in English. In M.P.R. van den Broecke & A. Cohen (Eds.), *Proceedings of the Tenth International Congress of Phonetic Sciences,* (pp. 529-536). Dordrecht, Cinnaminson: Foris.

Bartkova, K. (1991). Speaking rate modelization in French application to speech synthesis. *Proceedings of the Twelveth International Congress of Phonetic Sciences* (pp. 482-485). Aix en Provence, France.

Benguerel, A.-P. (1977). Velar coarticulation in French: A fiberscopic study. *Journal of Phonetics*, **5,** 145-158.

Bertelson, P., Vroomen, J., & de Gelder, B. (1997). Auditory-visual interaction in voice localization and in bimodal speech recognition: The effects of desynchronization. In C. Benoît & R. Campbell (Eds.), *Proceedings of the workshop on audio-visual speech processing: Cognitive and computational Approaches* (pp. 97-100). Rhodes, Greece: ESCA.

Bertinetto, P.-M. (1999). La sillabazione dei nessi /sC/ in italiano: un'eccezione alla tendenza "universale"? In P. Benincà, A. Mioni, & L. Vanelli (Eds.), *Fonologia e morfologia dell'italiano e dei dialetti d'Italia*, Atti del XXXI Congresso della SLI (pp. 71-96). Roma: Bulzoni.

Bertinetto, P.-M. (2004). On the undecidable syllabification of /sC/ clusters in Italian: Converging experimental evidence. *Italian Journal of Linguistics (Rivista di Linguistica)*, **16,** 349-372.

Bever, T.G. 1970. The cognitive basis for linguistic structures. In J. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.

Bonnot, J.F.P., Chevrie-Muller, C., Arabia-Guidet, C., Maton, B., & Grenier, G.F. (1986). Coarticulation and motor encoding of labiality and nasality in CVCVCV nonsense words. *Speech Communication*, **5,** 83-95.

Bowers, J.S., & Davis, C.J. (2004). Is speech perception modular or interactive? *Trends in Cognitive Sciences*, **8,** 3-5.

Bregman, A.S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Brent, M. (1999). Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences*, **3,** 294-301.

Browman, C., & Goldstein, L. (1995). Gestural syllable position effects in American English. In F. Bell-Berti & L.J. Raphael (Eds.), *Producing speech: A festschrift for Katherine Safford Harris* (pp. 19-34). Woodbury, N.Y.: American Institute of Physics Press.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996). Perception of lexical tone across languages: Evidence for a linguistic mode of processing. *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 2514-2516). Philadelphia.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production (VII). Development, writing and other language processes* (pp.257-294). London: Academic Press.

Byrd, D. (1994). *Articulatory timing in English consonant sequences.* PhD dissertation, UCLA. (publié sous *UCLA Working Papers in Phonetics*, **86**).

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

Cairns, P., Shillcock, R., Chater, N., & Levy, J.P. (1995). Bottom-up connectionist modeling of speech. In J.P. Levy, D. Bairaktaris, J.A. Bullinaria, & P. Cairns (Eds.), *Connectionist models of memory and language* (pp. 289-310). London: UCL Press.

Cairns, P., Shillcock, R., Chater, N., & Levy, J.P. (1997). Bootstrapping Word Boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, **33,** 111-153.

Caramazza, A., Laudanna, A., Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, **28,** 297-332.

Carlson, R., Granström, B., Lindblom, B., & Rapp, K. (1972). Some timing and fundamental frequency characteristics of Swedish sentences: Data, rules, and a perceptual evaluation. *STL-QPSR 4/1972* (pp. 11-19).

Chéreau, C., Gaskell, M.G., & Dumay, N. (sous presse). Reading spoken words: Orthographic effects in auditory priming. *Cognition*.

Ching, Y.C. (1985). Lipreading Cantonese with voice pitch. *Journal of the Acoustical Society of America*, **77,** Supplement 1, S39-S40. (Abstract)

Ching, Y.C. (1988). Voice pitch information for the deaf. *Proceedings of the First Asian-Pacific Regional Conference on Deafness* (pp. 340-343). Hong Kong.

Chodorow, M.S. (1979). Time-compressed speech and the study of lexical and syntactic processing. in W. Cooper & E. Walker (Eds.), *Sentence processing* (pp. 87-111). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Cholin, J., Schiller, N.O., & Levelt, W.J.M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, **50,** 47-61.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Christiansen, M.H., Allen, J., & Seidenberg, M.S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, **13,** 221-268.

Christie, W. (1974). Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America*, **55,** 819-821.

Christophe, A. (1993). *Rôle de la prosodie dans la segmentation en mots.* Thèse de doctorat non-publiée. EHESS, Paris.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, **95,** 1570-1580.

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

Clements, G.N., & Keyser, S.J. (1983). CV phonology: A generative theory of the syllable. *Linguistic Inquiry Monograph* **9**. Cambridge, MA: MIT Press.

Cohen J.D., MacWhinney B., Flatt M., & Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, **25,** 257-271.

Cohen, A. (1987). Juncture revisited. In R. Channon & L. Shockey (Eds.), *In honour of Ilse Lehiste* (pp. 7-18). Dordrecht: Foris.

Cole, R.A. (1973). Listening of mispronunciations: A measure of what we hear during speech, *Perception & Psychophysics*, **13,** 153-156.

Cole, R.A., & Jakimik, J. (1978). Understanding speech: How words are heard? In G. Underwood (Ed.), *Strategies of information processing* (pp. 67-116). London: Academic Press.

Cole, R.A., & Jakimik, J. (1980a). A model of speech perception. In R.A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133-164). Hillsdale, N.J.: Erlbaum.

Cole, R.A., & Jakimik, J. (1980b). How are syllables used to recognize words? *Journal of the Acoustical Society of America*, **67,** 965-970.

Cole, R.A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review*, **81,** 348-371.

Connine, C.M., & Clifton, C.E., (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, **2,** 291-299.

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

Content, A. Bagou, O., Frauenfelder, U.H., & Fougeron, C. (2004). Indices des frontières lexicales avec des pseudomots: Une étude préliminaire. *Actes des XXVèmes Journées d'Etudes sur la Parole* (pp. 137-140), Fès, Maroc.

Content, A., & Sternon, P. (1994). Modelling retroactive context effects in spoken word recognition with a simple recurrent network. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 207-212). Hillsdale, NJ: Erlbaum.

Content, A., Dumay, N., & Frauenfelder, U.H. (1999). Segmentation syllabique chez l'enfant. *Actes des IIèmes Journées d'Etudes Linguistiques* (pp. 69-74), Nantes, France.

Content, A., Kearns, R.K., & Frauenfelder, U.H. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language*, **45,** 177-199.

Content, A., Mousty, P., & Radeau, M. (1990). Brulex: Une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique*, **90,** 551-566.

Cooke, M.P., & Ellis, D.P.W. (2001). The auditory organization of speech and other sources in listeners and computational models. *Speech Communication*, **35,** 141-177.

Coursil, J. (1992). *Grammaire analytique du français contemporain. Essai d'intelligence artificielle et de linguistique générale.* Thèse de doctorat non-publiée, Université de Caen.

Crompton, A. (1982). Syllables and segments in speech production. In A. Cutler (Ed.), *Slip of the tongue and language production.* The Hague: Mouton.

Crystal, T.H., & House, A.S. (1988). The duration of American-English vowels: An overview. *Journal of Phonetics*, **16,** 263-284.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, **29,** 201-220.

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing* (pp.105-121). Cambridge, MA: MIT Press.

Cutler, A., & Butterfield, S. (1990). *Syllabic lengthening as a word boundary cue.* In R. Seidl (Ed.), *Proceedings of the 3rd Australian International Conference on Speech Science and*

Technology (pp. 324-328). Canberra: Australian Speech Science and Technology Association.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory & Language*, **31,** 218-236.

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., & Chen, H.-C. (1995). Phonological similarity effects in Cantonese word recognition. *Proceedings of the Thirteenth International of Phonetic Sciences* (pp. 106-109). Stockholm.

Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, **59**, 165–179.

Cutler, A., & Clifton, C.E. (1984). The use of prosodic information in word recognition. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 183-196). Hillsdale, N.J.: Erlbaum.

Cutler, A., & Koster, M. (2000). Stress and lexical activation in Dutch. *Proceedings of the Sixth International Conference on Spoken Language Processing* (pp. 593-596). Beijing.

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.

Cutler, A., & Norris, D. (2002). The role of strong syllables in segmentation for lexical access. In G.T.M. Altman (Ed.), *Psycholinguistics*, **Vol. 1** (pp. 157-177). London: Routledge (Critical concepts in psychology).

Cutler, A., & Otake, T. (1996). The processing of word prosody in Japanese. In P. McCormack & A. Russell (Eds.), *Proceedings of the Sixth Australian International Conference on Speech Science and Technology* (pp. 599-604). Canberra: Australian Speech Science and Technology Ass.

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America*, **105,** 1877-1888.

Cutler, A., & van Donselaar, W.A. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech*, **44,** 171-195.

Cutler, A., Dahan, D., & Van Donselaar, W.A. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, **40,** 141-202.

Cutler, A., Demuth, K., & McQueen, J.M. (2002). Universality versus language-specificity in listening to running speech. *Psychological Science*, **13,** 258-262.

Cutler, A., McQueen, J.M., Norris, D., & Somejuan, A. (2001). The roll of the silly ball. In E. Dupoux *et al.* (Eds.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler* (pp. 181-194). Cambridge, MA: MIT Press.

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

Cutting, J.E. (1975). Aspects of phonological fusion. *Journal of Experimental Psychology: Human Perception and Performance*, **1,** 105-120.

Dahan, D., & Tanenhaus, M.K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **30,** 498-513.

Dahan, D., Magnuson, J.S., Tanenhaus, M.K., & Hogan, E. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, **16,** 507-534.

Darwin, C.J. (1997). Auditory grouping. *Trends in Cognitive Sciences*, **1,** 327-333.

Davis, M.H., Marslen-Wilson, W.D., & Gaskell, M.G. (2002). Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **28,** 218-244.

Dejean de la Bâtie, B. (1993). *Word boundary ambiguity in spoken French.* Unpublished doctoral dissertation, Monash University, Victoria, Australia.

Delattre, P. (1966). *Studies in French and comparative phonetics.* The Hague: Mounton & Co.

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

Dell, F., & Vergnaud, J.-R. (1984). Les développements récents en phonologie: Quelques idées centrales. In F. Dell, D Hirst & J.-R. Vergnaud (Eds.), *Forme sonore du langage* (pp. 1-42). Paris: Hermann.

Dijkstra, T., Fieuws, S., & Roelofs, A. (1995). Orthographic effects on phoneme monitoring. *Canadian Journal of Experimental Psychology*, **49,** 264-271.

Dufour, S., & Peereman, R. (2003a). Lexical competition in phonological priming: Assessing the role of phonological match and mismatch lengths between primes and targets. *Memory & Cognition*, **31,** 1271-1283.

Dufour, S., & Peereman, R. (2003b). Inhibitory priming effects in auditory word recognition: When the target's competitors conflict with the prime word. *Cognition*, **88,** B33-B44.

Dumay, N., & Gaskell, M.G. (2005). Do words go to sleep? Exploring consolidation of spoken forms through direct and indirect measures. *Behavioral and Brain Sciences*, **28,** 69-70.

Dumay, N., & Gaskell, M.G. (sous presse). Sleep-associated changes in the mental representation of spoken words. *Psychological Science.*

Dumay, N., Benraïss, A., Barriol, B., Colin, C., Radeau, M., & Besson, M. (2001). Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words. *Journal of Cognitive Neuroscience*, **13,** 121-143.

Dumay, N., Frauenfelder, U.H., & Content, A. (2002). The role of the syllable in lexical segmentation in French: Word-spotting data. *Brain and Language*, **81,** 144-161.

Dumay, N., Gaskell, M. G., & Feng, X. (2004). A day in the life of a spoken word. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 339-344). Mahwah, NJ: Erlbaum.

Dupoux, E. (1993). The time course of prelexical processing: The syllable hypothesis revisited. In G.T.M. Altmann, & R. Shillcock (Eds.), *Cognitive models of speech processing: Second Sperlonga Meeting* (pp. 81-114). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

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

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

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

Encrevé, P. (1988). *La liaison avec et sans enchaînement.* Paris: Le Seuil.

Fant, G. (1962). Descriptive analysis of the acoustic aspects of speech. *Logos*, **5,** 3-17. (Publié également in Fant, G. (1973). *Speech sounds and features* (pp.17-23). Cambridge, MA: MIT Press).

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

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.

Fletcher, J. (1991). Rhythm and final lengthening in French. *Journal of Phonetics*, **19,** 193-212.

Floccia, C., Kolinsky, R., Dodane, C., & Morais, J. (2003). Discriminating spoken words in French: the role of the syllable and the CV phonological skeleton. *Language and Cognitive Processes*, **18,** 241-267.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Fónagy, I., & Magdics, K. (1960). Speed of utterance in phrases of different lengths. *Language and Speech*, **4,** 179-92.

Foss, D.G., & Gernsbacher, M.A. (1983). Cracking the dual code: Toward a unitary model of phonetic identification. *Journal of Verbal Learning and Verbal Behavior*, **22,** 609-632.

Fougeron, C. (1998). *Variations articulatoires en début de constituants prosodiques de différents niveaux en français.* Thèse de doctorat non-publiée. Université Paris III – Sorbonne Nouvelle, Paris, France.

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

Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, **101,** 3728-3740.

Fougeron, C., & Steriade, D. (1997) Does deletion of French schwa lead to neutralization of lexical distinctions? *Proceedings of the 5th European Conference on Speech Communication and Technology*, Vol. 7 (pp. 943-937). Rhodes, Greece: University of Patras, Wire Communications Laboratory.

Fougeron, C., Bagou, O., Stefanuto, M., & Frauenfelder, U.H. (2002). A la recherche d'indices de frontière lexicale dans la resyllabation. *Actes des XXIVème Journées d'Etude sur la Parole* (pp. 125-128), Nancy, France.

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

Fowler, C.A. (1990). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, **88,** 1236-1249.

Fowler, C.A., & Rosenblum, L. (1991). The perception of phonetic gestures. In I.G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 33-59). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Fox, R.A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 526-40.

Frauenfelder, U.H., & Content, A. (1999). The role of the syllable in spoken word recognition: Access or segmentation? *Actes des IIèmes Journées d'Etudes Linguistiques* (pp. 1-8), Nantes, France.

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

Frauenfelder, U.H., & Peeters, G. (1998). Simulating the time-course of word recognition: An analysis of lexical competition in TRACE. In J. Grainger & A. Jacobs (Eds.), *Localist*

connectionist approaches to human cognition (pp. 101-146). Hillsdale, NJ: Lawrence Erlbaum Associates.

Frauenfelder, U.H., & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1991* (pp. 165-183). Dordrecht: Kluwer.

Frauenfelder, U.H., Scholten, M., & Content, A. (2001). Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes*, **16,** 583-607.

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

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* Bloomington: Indiana University Linguistics Club.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 559-586). Hillsdale, N.J.: Erlbaum.

Frazier, L. (1989). Against lexical generation of syntax. In W.D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 505-528). Cambridge, MA: MIT Press.

Frazier, L., & Clifton, C.E. (1996). *Contrual.* Cambridge, MA: MIT Press.

Freedman, D. (1992). *Frequency effects in auditory word recognition in the word spotting-task.* Unpublished undergraduate project, Department of Experimental Psychology, University of Cambridge, UK.

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.

Fromkin, V. (1965). Some phonetic specifications of linguistic units: An electromyographic investigation. *UCLA Working Papers in Phonetics*, **3,** 1-184.

Fry, D.B., Abramson, A.S., Eimas, P.D., & Liberman, A.M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, **5,** 171-189.

Fujimura, O. (1990). Methods and goals of speech production research. *Language and Speech*, **33,** 195-258.

Fujimura, O., & Lovins, J.B. (1978). Syllables as concatenative phonetic units. In A. Bell & J.B. Hooper (Eds.), *Syllables and segments* (pp. 107-120). Amsterdam: North-Holland.

Gaitenby, J. (1965). The elastic word. *Haskins Laboratories: Status Report on Speech Research* **SR-2,** 3.1-3.12.

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

Garding, E. (1967). Internal juncture in Swedish. *Travaux de l'Institut de Phonétique de Lund*, **6**.

Gaskell, M.G., & Marslen-Wilson, W.D. (1997). Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes*, **12,** 613-656.

Gaskell, M.G., & Marslen-Wilson, W.D. (1999). Ambiguity, competition and blending in spoken word recognition. *Cognitive Science*, **23,** 439-462.

Gaskell, M.G., & Marslen-Wilson, W.D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology,* **45,** 220-266.

Gaskell, M.G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, **89,** 105-132.

Gaygen, D.E., & Luce, P.A., (2002). Troughs and bursts: Probabilistic phonotactics and lexical activation in the segmentation of spoken words in fluent speech. In C.T.M. McLennan, P.A. Luce, G. Mauner, & J. Charles-Luce (Eds.), *University at Buffalo Working Papers on Language and Perception*, **1,** 496-549.

Gimson, A. (1980). *An introduction to the pronunciation of English*, 3rd edition. London: Edward Arnold.

Goetry, V., & Kolinsky, R. (2000). The role of rhythmic cues for speech segmentation in monolingual and bilingual listeners. *Psychologica Belgica*, **40,** 115-152.

Goldinger, S.D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **22,** 1166-1183.

Goldinger, S.D. (1998). Echoes of echoes. *Psychological Review*, **105,** 251-279.

Goldinger, S.D. (2000). The role of perceptual episodes in lexical processing. In A. Cutler, J.M. McQueen, & R. Zondervan (Eds.), *Spoken word access processes* (pp. 155-158). Nijmegen, The Netherlands: Max-Planck Institute for Psycholinguistics.

Goldinger, S.D., Kleider, H.M., & Shelley E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory & Cognition*, **27,** 328-338.

Goldinger, S.D., Luce, P.A., & Pisoni, D.B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, **28,** 501-518.

Goldinger, S.D., Luce, P.A., & Pisoni, D.B., & Marcario, J.K. (1992). Form-based priming in spoken word recognition: the roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **18,** 1211-1238.

Goslin, J. Content, A., & Frauenfelder, U.H. (1999). Syllable segmentation: Are humans consistent? *Proceedings of Eurospeech' 99* (pp. 1683-1686), Budapest, Hungary.

Goslin, J., & Frauenfelder, U.H. (2000). A comparison of theoretical and human syllabification. *Language and Speech*, **44,** 409-436.

Gow, D.W. Jr., & Gordon, P.C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, **21,** 344-359.

Gow, D.W. Jr., & Gordon, P.C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, **21,** 344-359.

Grainger, J., Colé, P., & Seguí, J. (1991). Masked morphological priming in visual word recognition. *Journal of Memory and Language*, **30,** 370-384.

Grammont, M. (1933). *Traité de phonétique.* Paris: Delagrave.

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., Chéreau, C., & Seguí, J. (2000). Where is the /b/ in "absurde" [apsyrd]? It is in the French listeners' minds. *Journal of Memory and Language*, **43,** 618-639.

Hamburger, M.B., & Slowiaczek, L.M. (1996). Phonological priming reflects lexical competition in auditory word recognition. *Psychonomic Bulletin and Review*, **3,** 520-525.

Harrington, J., Laver, J., & Cutting, D. (1986). Word-structure reduction rules in automatic continuous speech recognition. In R. Lawrence (Ed.), *Proceedings of the Institute of Acoustics* (Vol. **8,** pp. 451-460). Edinburgh: Institute of Acoustics.

Harrington, J., Watson, G., & Cooper, M. (1989). Word boundary detection in broad class and phoneme strings. *Computer Speech and Language*, **3,** 367-382.

Harris, Z.S. (1954). Distributional Structure. *Word*, **10,** 146-162.

Harris, Z.S. (1955). From phoneme to morpheme. *Language*, **31,** 190-222.

Hoard, J.E. (1966). Juncture and syllable structure in English. *Phonetica*, **15,** 96-109.

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

Jackendoff, R. (1997). *The architecture of the language faculty.* Cambridge, MA: MIT Press.

Jackson, A., & Morton, J. (1984). Facilitation of auditory word recognition. *Memory & Cognition*, **12,** 568-574.

Jacobs, A.M., Rey, A., Ziegler, J.C, & Grainger, J. (1998). MROM-P: An interactive activation, multiple read-out model of orthographic and phonological processes in visual word recognition. In J. Grainger & A.M. Jacobs (Eds.), *Localist connectionist approaches to human cognition*. Mahwah, NJ.: Erlbaum.

Jones., D. (1931). The "word" as a phonetic entity. *Le maître phonétique, 3ième série,* **36,** 60-65.

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

Jusczyk, P.W., & Luce, P.A. (2002). Speech perception and spoken word recognition: Past and present. *Ear and Hearing*, **23,** 2-40.

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

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

Kager, R. (1989). *A metrical theory of stress and destressing in English and Dutch.* Dordrecht: Foris.

Kahn, D. (1976). *Syllable-based Generalizations in English Phonology*. PhD dissertation. Cambridge, MA: MIT Press.

Karlsson, F. (1983). *Suomen kielen äänne- ja muotorakenne* (Finnish phonological and morphological structure). Helsinki: WSOY.

Kaye, J., Lowenstamm, J., & Vergnaud, J.-R. (1990). Constituent structure and government in phonology. *Phonology*, **7,** 193-231.

Keating, P.A., & Wright, R. (1994). Effects of position-in-syllable on consonant articulation and acoustics. *Journal of the Acoustical Society of America*, **95,** 2819.

Keating, P.A., Wright, R., & Zhang, J. (1999). Word-level asymmetries in consonant articulation. *UCLA Working Papers in Phonetics*, **97,** 157-173.

Kessinger, R.H., & Blumstein, S.E. (1998). Effects of speaking rate on voice onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, **26,** 117-128.

Klatt, D.H. (1974). The duration of [s] in English words. *Journal of Speech and Hearing Research*, **17,** 51-63.

Klatt, D.H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59,** 1208-1221.

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

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.

Kohler, K.J. (1976). Die Instabilität wortfinaler Alveolarplosive im Deutschen: Eine elektropalatographische Untersuchung. *Phonetica*, **33,** 1-30.

Kohler, K.J. (1994). Glottal stops and glottalization in German. *Phonetica*, **51,** 38-51.

Kohler, K.J., & Hartcastle, W. (1974). The instability of final alveolars in English and German: Proposal for an instrument investigation. *Speech Communication Seminar*, **2,** 95-98. Stockholm, Sweden.

Koster, M., & Cutler, A. (1997). Segmental and suprasegmental contributions to spoken-word recognition in Dutch. *Proceedings of the Fifth European Conference on Speech Communication and Technology* (pp. 2167-2170). Rhodes, Greece: University of Patras, Wire Communications Laboratory.

Krakow, R.A. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velic gestures.* Unpublished PhD dissertation, Yale University.

Krakow, R.A. (1993). Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate. In M.K. Uffman & R.A. Krakow (Eds.), *Nasals, nasalisation and the velum* (pp. 87-116). San Diego: Academic Press.

Krakow, R.A. (1999). Physiological organization of syllables: a review. *Journal of Phonetics*, **27,** 23-54.

Lahiri, A., & Marslen-Wilson, W.D. (1991). The mental representation of lexical form: A phonological approach to the mental lexicon. *Cognition*, **38,** 245-294.

Landauer, T.K., & Streeter, L.A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of verbal learning and verbal behavior*, **12,** 119-131.

Laporte, E. (1993). Phonetic syllables in French: Combinations, structure, and formal definitions. *Acta Linguistica Hungarica*, **41,** 175-189.

Lehiste, I. (1960). An acoustic-phonetic study on internal open juncture. *Phonetica*, **5:** suppl.

Lehiste, I. (1961). Acoustic studies of boundary signals. *Proceedings of the Fifth International Congress of Phonetic Sciences* (pp. 178-187). Helsinki, Finland.

Lehiste, I. (1964). Juncture. In E. Zwirner & W. Bethge (Eds.), *Proceedings of the Fifth International Congress of Phonetic Sciences*, Utrecht 1964 (pp. 172-200). Basel – New York: Karger.

Lehiste, I. (1972). The timing of utterance and linguistic boundaries. *Journal of the Acoustical Society of America*, **51,** 2018-2024.

Léon, P. (1966). *Prononciation du français standard.* Paris: Didier.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W.J.M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, **42,** 1-22.

Levelt, W.J.M. (1999). Models of word production. *Trends in Cognitive Sciences*, **3,** 223-232.

Levelt, W.J.M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, **50,** 239-269.

Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, **22,** 1-72.

Liberman, A.M. (1970). The grammars of speech and language. *Cognitive Psychology*, **1,** 301-323.

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

Liberman, A.M., & Mattingly, I.G. (1989). A specialization for speech perception. *Science*, **243,** 489-494.

Liberman, A.M., & Whalen, D.H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, **4,** 187-196.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, **74,** 431-459.

Liberman, A.M., Delattre, P., Cooper, J. Gertsman, L. (1954). The role of consonant-vowel transitions in the perception of stops and nasal consonants. *Psychological Monographs*, **68,** 1-13.

Liberman, M.Y., & Streeter, L.A. (1978). Use of nonsense-syllable mimicry in the study of prosodic phenomena. *Journal of the Acoustical Society of America*, **63,** 231-233.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. *Papers in Linguistics, University of Stockholm*, **21,** 1-59.

Luce, P.A. (1986a). *Neighborhoods of words in the mental lexicon*. Unpublished doctoral dissertation, Indiana University at Bloomington, Indiana.

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

Luce, P.A. (1989). *Similarity neighborhoods of words in the mental lexicon: A computational analysis.* Bloomington, IN: Speech Research Laboratory, Indiana University.

Luce, P.A., & Cluff, M.S. (1998). Delayed commitment in spoken word recognition: Evidence from cross-modal priming. *Perception & Psychophysics*, **60,** 484-490.

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

Luce, P.A., Goldinger, S.D., Auer E.T., & Vitevitch, M.S. (2000). Phonetic priming, Neighborhood activation, and PARSYN. *Perception & Psychophysics*, **62,** 615-625.

Luce, P.A., Pisoni, D.B., & Goldinger, S.D. (1990). Similarity neighborhoods of spoken words. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 122-147). Cambridge, MA: MIT Press.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge, MA: Cambridge University Press.

Magnuson, J.S., McMurray, B., Tanenhaus, M.K., & Aslin, R. (2003). Lexical effects on compensation for coarticulation: The ghost of Christmash past. *Cognitive Science*, **27,** 285-298.

Magnuson, J.S., Tanenhaus, M.K., Aslin, R., & Dahan, D. (2003). The microstructure of spoken word recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, **132,** 202-227.

Malécot, A., Johnston, R., & Kizziar, P.A. (1972). Syllabic rate and utterance length in French. *Phonetica*, **26,** 235-251.

Manaster-Ramer, A. (1996). A letter from an incompletely neutral phonologist. *Journal of Phonetics*, **24,** 477-489.

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.

Marcus, S.M. (1981). ERIS–Context-sensitive coding in speech perception. *Journal of Phonetics*, **9,** 197-220.

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

Marslen-Wilson, W.D. (1993). Issues of process and representation in lexical access. In G.T.M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: Second Sperlonga Meeting* (pp. 187-210). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Marslen-Wilson, W.D., & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition*, **8,** 1-71.

Marslen-Wilson, W.D., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, **101,** 653-675.

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

Marslen-Wilson, W.D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, **15**, 576-585.

Marslen-Wilson, W.D., Ford, M., Older, L., & Zhou, X. (1996). The combinatorial lexicon: Priming derivational affixes. In G. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 223-227). Mahwah, NJ: Erlbaum.

Marslen-Wilson, W.D., Tyler, L.K., & Seidenberg, M.S. (1978). Sentence processing and the clause boundary. In W.J.M. Levelt & G. Flores D'arcais (Eds.), *Studies in the Perception of Language.* London: Wiley.

Marslen-Wilson, W.D., Tyler, L.K., Waksler, S., & Older L. (1994). Morphology and meaning
in the mental lexicon. *Psychological Review*, **101,** 3-33.

Martinet, A., (1945). *La prononciation du français contemporain.* Genève: Droz.

Matter, J.F. (1986). *A la recherche des frontières perdues: Etude sur la perception de la parole en français.* Amsterdam: De Werelt.

Mattys, S. L., Pleydell-Pearce, C. W., Melhorn, J. F., & Whitecross, S. E. (2005). Detecting silent pauses in speech: A new tool for measuring on-line lexical and semantic processing. *Psychological Science*, **16,** 958-964.

Mattys, S.L., & Samuel, A.G. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory and Language*, **36,** 87-116.

Mattys, S.L., & Samuel, A.G. (2000). Implications of stress pattern differences in spoken word recognition. *Journal of Memory and Language*, **42,** 571-596.

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

Mattys, S.L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics*, **62,** 253-265.

Mattys, S.L. (2004). Stress versus coarticulation: Towards an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, **30,** 397-408.

Mattys, S.L., & Clark, J.H. (2002). Lexical activity in speech processing: evidence from pause detection. *Journal of Memory and Language*, **47,** 343-359.

Mattys, S.L., & Jusczyk, P.W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, **78,** 91-121.

Mattys, S.L., & Samuel, A.G. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory and Language*, **36,** 87-116.

Mattys, S.L., Jusczyk, P.W., Luce, P.A., & Morgan, J.L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, **38,** 465-494.

Mattys, S.L., White, L., & Melhorn, J.F (2005). Integration of multiple segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General,* **134,** 477-500.

McClelland, J. L. and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, **88,** 375-407.

McClelland, J.L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, **86,** 287-330.

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

McDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, **101,** 676-703.

McMurray, B., & Aslin, R.N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, **95,** B15-B26.

McMurray, B., Tanenhaus, M.K., & Aslin, R.N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, **86,** B33-B42.

McQueen, J.M. (1996). Word-spotting. In F. Grosjean & U.H. Frauenfelder (Eds.), *Spoken word recognition paradigms.* A special issue of *Language and Cognitive Processes* (Vol. **11,** pp. 695-699).

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

McQueen, J.M. (2003). The ghost of Christmas future: Didn't Scrooge learn to be good? Commentary on Magnuson, McMurray, Tanenhaus, & Aslin (2003). *Cognitive Science*, **27,** 795-799.

McQueen, J.M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes*, **10,** 309-331.

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.

McQueen, J.M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, **25,** 1363-1389.

McQueen, J.M., Otake, T., & Cutler, A. (2001). Rhythmic Cues and Possible-Word Constraints in Japanese Speech Segmentation. *Journal of Memory and Language*, **45,** 103-132.

Mehler, J. (1981). The role of syllables in speech processing: infant and adult data. *Philosophical Transactions of the Royal Society, Series B*, **295,** 333-352.

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

Mehler, J., Dupoux, E., & Seguí, J. (1990). Constraining models of lexical access: The onset of word recognition. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 236-262). Cambridge, MA: MIT Press.

Menzerath, P., & de Lacerda, A. (1933). *Koartikulation steuerung und lautabgrenzung.* Bonn: Ferdinand Dümmlers Verlag.

Meunier, C., Frauenfelder, U.H., & Content, A. (1996) La segmentation lexicale: Contribution des indices segmentaux et suprasegmentaux. *Actes des XXIe Journées d'Etude sur la Parole* (pp. 31-34). Avignon, France.

Miller, J.L., & Dexter, E.R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, **14,** 369-378.

Mills, C.B. (1980). Effects of match between listener expectancies and coarticulatory cues on the perception of speech. *Journal of Experimental Psychology: Human Perception and Performance*, **6,** 528-535.

Minematsu, N., & Hirose, K. (1995). Role of prosodic features in the human process of perceiving spoken words and sentences in Japanese. *Journal of the Acoustical Society of Japan*, **16**, 311–320.

Mirman, D., McClelland, J.L., & Holt, L.L. (2005). Computational and behavioral investigations of lexically induced delays in phoneme recognition. *Journal of Memory and Language*, 52, 424-443.

Monsen, R.B., & Engebretson, A.M. (1977). Study of variations in the male and female glottal wave. *Journal of the Acoustical Society of America*, **62,** 981-993.

Morais J., Castro S.L., & Kolinsky, R. (1991). La reconnaissance des mots chez les illettrés. In R. Kolinsky, J. Morais, & J. Seguí (Eds.), *La reconnaissance des mots dans les différentes modalités sensorielles: Etudes de psycholinguistique expérimentale* (pp. 59-80). Presses Universitaires de France.

Morais, J. (1985). Literacy and awareness of the units of speech: implications for research on the units of perception. *Linguistics*, **23,** 707-721.

Morais, J., & Kolinsky, R. (2001). The literate mind and the universal human mind. In E. Dupoux *et al.* (Eds.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler* (pp. 463-480). Cambridge, MA: MIT Press.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, **24,** 45-64.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, **7,** 323-331.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, **76,** 165-178.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P.A. Kolers, M.E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language.* New York: Plenum.

Munhall, K.G., Gribble, P., Sacco, L., & Ward, M. (1999). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, **58,** 351-362.

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

Nakatani, L.H., & Schaffer, J.A. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America*, **63,** 234-245.

Nespor, M. (1977). Segment and word durational correlates of syntactic boundary in Italian. In R. Rodman (Ed.), *Explorations in language and linguistics*. University of North Carolina: Chapel Hill.

Nooteboom, S.G. (1972). *Production and perception of vowel duration: A study of durational properties of vowels in Duch.* Unpublished PhD dissertation, University of Utrecht, Utrecht, The Netherlands.

Nooteboom, S.G. (1981). Lexical retrieval from fragments of spoken words: Beginnings vs endings. *Journal of Phonetics*, **9,** 407-424.

Norris, D. (1990). A dynamic-net model of human speech recognition. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 122-147). Cambridge, MA: MIT Press.

Norris, D. (1993). Bottom-up connectionist models of "interaction". In G.T.M. Altmann & R. Shillcock (Eds.), Cognitive models of speech processing: Second Sperlonga Meeting (pp. 211-234). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

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

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., McQueen, J.M., Cutler, A., Butterfield, S., & Kearns, R. (2001). Language-universal constraints on speech segmentation. *Language and Cognitive Processes*, **16,** 637-660.

Noske, R. (1982). Syllabification and syllable changing rules in French. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations*, Vol. 2 (pp. 257-310). Dordrecht: Foris.

Nyrop, K. (1924). *Manuel phonétique du français parlé.* Paris: Editions Picard.

O'Connor, J.D., & Tooley, O.M. (1964). The perceptibility of certain word-boundaries. In X. Abercrombie, X. Fry, X. McCarthy, X. Scott, & J.L.M. Trim (Eds.), *In honour of Daniel Jones* (pp. 171-176). London: Longmans.

Oldfield, R.C. (1966). Things, words, and the brain. *Quarterly Journal of Experimental Psychology*, **18,** 340-353.

Oller, D.K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, **54,** 1235-1247.

Otake, T., Yoneyama, K., Cutler, A., & Van der Lugt, A.H. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America, 100* (6), 3831-3842.

Pasdeloup, V. (1990). *Modèle de règles rythmiques du français appliquées à la synthèse de la parole.* Thèse de doctorat non-publiée. Université de Provence.

Passy, P. (1917). *Les sons du français.* Paris: Didier

Peereman, R. (1999). Syllabification: Solution basée sur l'échelle de consonanticité. Rapport interne de l'Université de Bourgogne.

Peña, M., Bonatti, L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, **298,** 604-607.

Peterson, G.E., & Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, **24,** 175-184.

Pickett, J.M., & Pollack, I. (1963). Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech*, **6,** 151-164.

Pisoni, D.B., & Luce, P.A. (1987). Acoustic-phonetic representations in word recognition. *Cognition, **25**,* 21-52.

Pitt, M.A. (1995). The locus of lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **21,** 1037-1052.

Pitt, M.A., & McQueen, J.M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, **39,** 347-370.

Pitt, M.A., & Samuel, A.G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology*, **29,** 149-188.

Pitt, M.A., & Samuel, A.G. (sous presse). Word length and lexical activation: Longer is better. *Journal of Experimental Psychology: Human Perception and Performance.*

Pollack, I., & Pickett, J.M. (1963). The intelligibility of excerpts from conversation. *Language and Speech*, **6,** 165-171.

Pollack, I., & Pickett, J.M. (1964). Intelligibility of excerpts from fluent speech: Auditory vs. structural context. *Journal of Verbal Learning & Verbal Behavior*, **3,** 79-84.

Port, R.F. (1981). Linguistic timing factors in combination. *Journal of the Acoustical Society of America*, **69,** 262-274.

Port, R.F. (1996). The discreteness of phonetic elements and formal linguistics: A response to A. Manaster-Ramer. *Journal of Phonetics*, **24,** 491-511.

Pulgram, E. (1970). *Syllable, Word, Nexus, Cursus.* La Haye/Paris: Mouton.

Pustejovsky, J. (1995). *The generative lexicon.* Cambridge, MA: MIT Press.

Pylyshyn, Z.W. (1984). *Computation and cognition: Towards a foundation for cognitive science.* Cambridge, MA: MIT Press.

Quené, H. (1989). *The influence of acoustic-phonetic word boundary markers on perceived word segmentation in Dutch.* Unpublished doctoral dissertation, Rijksuniversiteit, Utrecht, The Netherlands.

Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, **20,** 331-350.

Quené, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustical Society of America*, **94,** 2027-2035.

Quené, H., & Koster, M.L. (1998). Metrical segmentation in Dutch: vowel quality or stress? *Language and Speech*, **41,** 185-202.

Quené, H., & Smith, Y. (1992). On the absence of word segmentation at "weark" syllables. *Proceedings of the International Conference on Spoken Language Processing* (Vol. 1, pp. 213-215). University of Alberta, Alberta.

Radeau, M., & Morais, J. (1990). The uniqueness point effect in the shadowing of spoken words. *Speed Communication*, **9,** 155-164.

Radeau, M., Morais, J., & Seguí, J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, **21,** 1297-1311.

Radeau, M., Morais, J., Mousty, P., & Bertelson, P. (2000). The effect of speaking rate on the role of the uniqueness point in spoken word recognition. *Journal of Memory and Language*, **42,** 406-422.

Radeau, M., Mousty, P., & Bertelson, P. (1989). The effect of the uniqueness point in spoken-word recognition. *Psychological Research*, **51,** 123-128.

Read, C.A., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic reading. *Cognition*, **24,** 31-44.

Reddy, R. (1976). Speech recognition by machines: A review. *Proceedings of the IEEE*, **64,** 501-531.

Remez, R.E. (2003). Establishing and maintaining perceptual coherence: Unimodal and multimodal evidence. *Journal of Phonetics*, **31,** 293-304.

Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S., & Lang, J.M. (1994). On the perceptual organization of speech. *Psychological Review*, **12,** 24-29.

Remez, R.E., Yang, C.Y., Piorkowski, R.L., Wissig, S., Batchelder, A., & Nam, H. (2002). The effect of variation in naturalness on phonetic perceptual identification. *Journal of the Acoustical Society of America*, **111,** 2432 (Abstract).

Repp, B.H., & Lin, H.-B. (1990). Integration of segmental and tonal information in speech perception. *Journal of Phonetics*, **18,** 481-495.

Repp, B.H., & Mann, V.A. (1981). Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America*, **69,** 1154-1163.

Repp, B.H., & Mann, V.A. (1982). Fricative-stop coarticulation: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **71,** 1562-1567.

Rialland, A. (1986). Schwa et syllabes en français. In L. Wetzels & E. Sezer (Eds.), *Studies in compensatory lengthening* (pp. 186-226). Dordrecht: Foris.

Rialland, A. (1994). The phonology and phonetics of extrasyllabicity in French. In P. Keating (Ed.), *Phonological structure and phonetic form. Papers in Laboratory Phonology III.* Cambridge, MA: University Press.

Rietveld, A.C.M. (1980). Word boundaries in the French language. *Language and Speech*, **23,** 289-296.

Rosenblum, L.D. (2005). The primacy of multimodal speech perception. In D.B. Pisoni & R.E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 51-78). Oxford: Blackwell Publishers.

Rousselot, P.J. (1901). *Principes de phonétique expérimentale.* Paris: Welter.

Rubin, P., Turvey, M.T., & van Gelder, P.V. (1976). Initial phonemes are detected faster in spoken words than in nonwords. *Perception & Psychophysics*, **19,** 394-398.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month old infants. *Science*, **274,** 1926-1928.

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.

Saffran, J.R., Newport, E.L., Aslin, R.N., Tunick, R.A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, **8,** 101-105.

Salverda, A.-P., Dahan, D., & McQueen, J.M. (2004). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, **90,** 51-89.

Samuel, A.G. (1981). Phonetic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, **110,** 474-494.

Samuel, A.G. (1989). Insights from a failure of selective adaptation: Syllable-initial and syllable-final consonants are different. *Perception & Psychophysics*, **45,** 485-493.

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. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, **12,** 348-351.

Samuel, A.G., & Pitt, M.A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, **48,** 416-434.

Sapir, E. (1933). La réalité psychologique des phonèmes. *Journal de Psychologie Normale et Pathologique*, **30,** 247-265.

Schaegis, A-L., Spinelli, E., & Welby, P. (2005). Perception of phonemically ambiguous spoken sequences in French. In B.G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (pp. 1937-1942). Mahwah, NJ: Erlbaum.

Schouten M.E.H., & Van Hessen, A.J. (1992). Modeling phoneme perception. I: Categorical perception. *Journal of the Acoustical Society of America*, **92,** 1841-1855.

Schriefers, H. (1999). Morphology and word recognition. In A.D. Friederici (Ed.), *Language comprehension: A biological perspective* (pp. 101-132). Berlin-Heidelberg: Springer-Verlag.

Schriefers, H., Zwitserlood, P., & Roelofs, A. (1991). The identification of morphologically complex spoken words: Continuous processing or decomposition? *Journal of Memory and Language*, **30,** 26–47.

Sebastián-Gallés, N., Dupoux, E., Seguí, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language*, **31,** 18-32.

Seguí, J. (1984). The syllable: A basic perceptual unit in speech processing. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 165-181). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Seguí, J., Dupoux, E., & Mehler, J. (1990). The role of the syllable in speech segmentation, phoneme identification, and lexical access. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 263-280). Cambridge, MA: MIT Press.

Seidenberg, M.S., & Tanenhaus, M.K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, **5,** 546-554.

Seidenberg, M.S., Tanenhaus, M.K., Leiman, J.L., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, **14,** 470-519.

Sekiguchi, T., & Nakajima, Y. (1999). The Use of Lexical Prosody for Lexical Access of the Japanese Language. *Journal of Psycholinguistics Research*, **28,** 439-454.

Selkirk, E. (1982). Syllables. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (Part II) (pp. 337-383). Dordrecht: Foris.

Shatzman, K.B., & McQueen, J. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics*, **68,** 1-16.

Shillcock, R. (1990). Lexical hypotheses in continuous speech. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 24-49). Cambridge, MA: MIT Press.

Shillcock, R., Bard, E.G., & Spensley, F. (1988). Some prosodic effects on human word recognition in continuous speech. *Proceedings of Speech '88, 7th FASE Symposium*, 827-834.

Shillcock, R., Lindsey, G., Levy, J.P., & Chater, N. (1992). A phonologically motivated input representation for the modelling of auditory word perception in continuous speech. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 408-413). Mahwah, NJ: Erlbaum.

Shillcock, R.C., & Bard, H.G. (1993). Modularity and the processing of closed-class words. In G.T.M. Altmann & R. Shillcock (Eds.), Cognitive models of speech processing: Second Sperlonga Meeting (pp. 163-185). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Slowiaczek, L. M., & Hamburger, M. B. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **18,** 1239-1250.

Slowiaczek, L.M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech*, **33,** 47-68.

Slowiaczek, L.M. (1991). Stress and context in auditory word recognition. *Journal of Psycholinguistic Research*, **20,** 465-81.

Slowiaczek, L.M., McQueen, J.M., Soltano, E., & Lynch, M. (2000). Phonological representations in prelexical speech processing: Evidence from form-based priming. *Journal of Memory and Language*, **43,** 530-560.

Slowiaczek, L.M., Nusbaum, H.C., & Pisoni, D.B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **13,** 64-75.

Smith, Z.M., Delgutte, B., & Oxenham, A.J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, **416,** 87-90.

Soto, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, **45,** 412-432.

Spinelli, E., McQueen, J.M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, **48,** 233-254.

Spivey-Knowlton, M.J., & Sedivy, J. (1995) Resolving attachment ambiguities with multiple constraints. *Cognition*, **55,** 227-267.

Stevens, K.N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, **111,** 1872-1890.

Stevens, K.N., & Blumstein, S.E. (1981). The search for invariant acoustic correlates of phonetic features. In P. Eimas & J.L. Miller (Eds.), *Perspectives on the study of speech* (pp. 1-38). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Straka, G. (1963). La division des sons du langage en voyelles et consonnes peut-elle être justifiée? *Travaux de Linguistique et Littérature, Université de Strasbourg*, **2,** 17-99.

Streeter, L., & Nigro, G. (1979). The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America*, **65,** 1533-1541.

Studdert-Kennedy, M. (1976). Speech perception. In N.J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 201-285). New York: Academic Press.

Suomi, K., McQueen, J.M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, **36,** 422-444.

Sweet, H. (1877). *Handbook of Phonetics.* Oxford: Clarendon. (Cité par Wood, S.A.J. (1993). Crosslinguistics cineradiographic studies of the temporal coordination of speech gestures. *Working papers*, **40,** Lund University, 251-263.)

Swinney, D. (1979). Lexical activation during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, **18,** 645-660.

Tabossi, P., Burani, C., & Scott, D. (1995). Word identification in fluent speech. *Journal of Memory and Language*, **34,** 440-467.

Taft, M., & Forster, K.I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, **14,** 638-647.

Taft, M., & Hambly, G. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language*, **24,** 320-335.

Taft, M., & Hambly, G. (1986). Exploring the Cohort model of word recognition. *Cognition*, **22,** 259-282.

Taft, M., Hambly, G., & Kinoshita, S. (1986). Visual and auditory recognition of prefixed words. *Quarterly Journal of Experimental Psychology*, **38,** 351-365.

Tanenhaus, M.K., Boland, J.E., Mauner, G., & Carlson, G.N. (1993). More on combinatory lexical information: Thematic structure in parsing and interpretation. In G.T.M. Altmann & R. Shillcock (Eds.), Cognitive models of speech processing: Second Sperlonga Meeting (pp. 297-319). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Tanenhaus, M.K., Garnsey, S., & Boland, J. (1990). Combinatory lexical information and language comprehension. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 383-408). Cambridge, MA: MIT Press.

Tanenhaus, M.K., Leiman, J.L., & Seidenberg, M.S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, **18,** 427-440.

Tranel, B. (1987). *The sounds of French: An introduction.* Cambridge: Cambridge University Press.

Trask, R.L. (1996). *A dictionary of phonetics and phonology.* London: Routledge.

Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language*, **27,** 87-104.

Treiman, R., Gross, J., & Cwikiel-Glavin, A. (1992). The syllabification of /s/ clusters in English. *Journal of Phonetics*, **20,** 383-402.

Trubetzkoy, N.S. (1939). Grundzüge der Phonologie. *Travaux du Cercle Linguistique de Prague*, **7**. (Pour une traduction française, voir Troubetzkoy, N.S. (1949). *Principes de Phonologie*. Paris: Klincksieck.)

Trueswell, J.C., & Tanenhaus, M.K. (1994). Towards a lexicalist framework of constraint-based syntactic ambiguity resolution. In C.E. Clifton & L. Frazier (Eds.), *Perspectives on sentence processing.* Hillsdale, N.J.: Erlbaum.

Umeda, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America*, **58,** 434-445.

Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America*, **61,** 846-858.

Umeda, N. (1978). Occurrence of glotal stops in fluent speech. *Journal of the Acoustical Society of America*, **64,** 88-94.

Umeda, N., & Coker, C.H. (1975). Subphonemic details in American English. In G. Fant & M.A.A. Tatham (Eds.), *Auditory analysis and perception* (pp. 539-564). London: Academic Press.

Vaissière, J. (1977). Premiers essais d'utilisation de la durée pour la segmentation en mots dans un système de reconnaissance. *Actes des VIIIèmes Journées d'Etudes sur la Parole* (pp. 345-352), Aix-en-Provence, France.

Vaissière, J. (1983). Language-independent prosodic features. In A Cutler & R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 53-66). Berlin: Springer-Verlag.

Vaissière, J. (1989). *Contribution à l'analyse des phénomènes de parole continue lue.* Thèse d'habilitation non-publiée. Université de Strasbourg, France.

Vaissière, J. (1992). Rhythm, accentuation and final lengthening in French. In J. Sundberg, L. Nord & R. Carlson (Eds.), *Music, language, speech and brain* (pp. 108-120). Stockholm: Wenner-Gren.

van der Hulst, H. (1984). *Syllable structure and stress in Dutch.* Dordrecht: Foris.

van der Lugt, A.H. (2001). The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics*, **63,** 811-823.

van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology*, **58A,** 251-273.

van Leyden, K., & van Heuven, V.J., (1996). Lexical stress and spoken word recognition: Dutch vs. English. In C. Cremers & M. den Dikken (Eds.), *Linguistics in the Netherlands 1996* (pp. 159-170). Amsterdam: John Benjamins.

Ventura, P., Kolinsky, R., Brito-Mendes, C., & Morais, J. (2001). Mental representations of the syllable internal structure are influenced by orthography. *Language and Cognitive Processes*, **16,** 393-418.

Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive processes*, **19,** 57-95.

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. (1997a). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **23,** 710-720.

Vroomen, J., & de Gelder, B. (1997b). Trochaic rhythm in speech segmentation. *Abstracts of the 38th Annual Meeting of the Psychonomic Society* (Vol. **2**, p. 73), Philadelphia, PA.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, **38,** 133-149.

Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: evidence from juncture misperceptions and word spotting. *Memory & Cognition*, **24,** 744-755.

Waals, J. (1999). *An experimental view of the Dutch syllable.* Doctoral dissertation, Utrecht University (*LOT Dissertation Series*, Vol. **18**). Utrecht: LOT.

Walsh Dickey, L. (1996). Limiting-domains in lexical access: Processing of lexical prosody. In M. Dickey & S. Tunstall (Eds.), *University of Massachusetts Occasional Papers in Linguistics, 19: Linguistics in the Laboratory* (pp. 133-155).

Wauquier-Gravelines, S. (1996). *Organisation phonologique et traitement de la parole continue: Contraintes prosodiques et phonologiques de l'accès lexical.* Thèse de doctorat non-publiée, Université Paris VII - Denis Diderot, Paris, France.

Weber, A. (2000). The role of phonotactics in the segmentation of native and non-native continuous speech. In A. Cutler, J.M. McQueen, & R. Zondervan (Eds.), *Spoken word access processes* (pp. 143-146). Nijmegen, The Netherlands: Max-Planck Institute for Psycholinguistics.

Wenk, B.J., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, **10,** 193-216.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II. *Psychologishe Forschung*, **4,** 301-350.

Whalen, D. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics*, **35,** 49-64.

Whalen, D. (1991). Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics*, **50,** 351-360.

Wheeldon, L.R., & Lahiri, A. (1997). Prosodic units in language production. *Journal of Memory and Language*, **37,** 356-381.

Wheeldon, L.R., & Lahiri, A. (2002). The minimal unit of phonological encoding: Prosodic or lexical. *Cognition*, **85,** B31-B41.

Wurm, L.H., & Samuel, A.G. (1997). Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *Journal of Memory and Language*, **36,** 165-187.

Ziegler, J.C., & Ferrand, L. (1998). Orthography shapes the perception of speech: the consistency effect in auditory word recognition. *Psychonomic Bulletin & Review*, **5,** 683-689.

Ziegler, J.C., Ferrand, L., & Montant, M. (2004). Visual phonology: The effects of orthographic consistency on different auditory word recognition tasks. *Memory & Cognition*, **32,** 732-741.

Ziegler, J.C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, **48,** 779-793.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken word processing. *Cognition*, **32,** 25-64.

Table des appendices

Appendice 1: Dumay, N., Frauenfelder, U.H., & Content, A. (2002). The role of the syllable in lexical segmentation in French: Word-spotting data. *Brain and Language*, 81, 144-161. 193
Appendice 2.1: Porteurs de l'Expérience 1 (et stimuli des Expériences 2 et 3) et énoncés d'origine. 230
Appendice 2.2: Remplisseurs de l'Expérience 1 et énoncés d'origine. 232
Appendice 3: Distributions des TRs par condition pour les Expériences 1, 4, 7 et 8. 234
Appendice 4.1: Porteurs de l'Expérience 4 (et stimuli de l'Expérience 5) et énoncés d'origine. 235
Appendice 4.2: Remplisseurs de l'Expérience 4 et énoncés d'origine. 237
Appendice 5.1: Enoncés de deux mots et phrases porteuses utilisés dans l'Expérience 6. 239
Appendice 5.2: Enoncés de deux mots et phrases porteuses utilisés dans l'Expérience 6b. 243
Appendice 6.1: Enoncés dont sont extraits les porteurs des Expériences 7 et 8. 245
Appendice 6.2: Enoncés dont sont extraits les remplisseurs de l'Expérience 7. 247
Appendice 6.3: Enoncés dont sont extraits les remplisseurs de l'Expérience 8. 248
Appendice 7: Test de fiabilité de la mesure du temps de réaction par le logiciel PsyScope (Cohen *et al.*, 1993). 249

Appendice 1: Dumay, N., Frauenfelder, U.H., & Content, A. (2002). The role of the syllable in lexical segmentation in French: Word-spotting data. *Brain and Language*, 81, 144-161.

THE ROLE OF THE SYLLABLE IN LEXICAL SEGMENTATION IN FRENCH: WORD-SPOTTING DATA

Nicolas Dumay [1,2], Uli H. Frauenfelder [3] & Alain Content [1,3]

[1] *Laboratoire de Psychologie Expérimentale, Université libre de Bruxelles, Belgium*

[2] *National Fund for Scientific Research, Belgium*

[3] *Laboratoire de Psycholinguistique Expérimentale, Université de Genève, Switzerland*

Running Title: LEXICAL SEGMENTATION IN FRENCH

Correspondence to:

Nicolas Dumay
Psycholinguistics Research Group
Department of Psychology
University of York
Heslington, York YO10 5DD
United Kingdom

E-mail: n.dumay@psych.york.ac.uk; Fax: 44 1904 433181

Abstract

Three word-spotting experiments assessed the role of syllable onsets and offsets in lexical segmentation. Participants detected CVC words embedded initially or finally in bisyllabic nonwords with aligned (CVC.CVC) or misaligned (CV.CCVC) syllabic structure. A misalignment between word and syllable onsets (Experiment 1) produced a greater perceptual cost than a misalignment between word and syllable offsets (Experiments 2 and 3). These results suggest that listeners rely on syllable onsets to locate the beginning of words. The implications for theories of lexical access in continuous speech are discussed.

Key words: syllable; spoken word recognition; lexical segmentation; word-spotting

The intuition that speech comes in syllables is probably universal and as ancient as the very beginnings of human thinking about language. Even though the introduction of the notion into phonology is relatively recent, the syllable is now generally considered a fundamental unit in modern accounts of many aspects of language behavior. Yet, somewhat paradoxically, despite the large amount of research that the syllable has generated and its ubiquitous appearance in linguistic and psycholinguistic theorizing, the exact role of syllables in the perception of speech and word recognition in continuous speech is still far from clear.

Early research into the role of the syllable in speech recognition focused on the question of its psychological reality as a "perceptual unit" (e.g. Savin & Bever, 1970). However, it has since appeared that this question was too vague and needed to be refined. One step in this direction was to distinguish between two fundamental functions of the perceptual mechanisms involved in speech processing (e.g. Norris & Cutler, 1985). One of these functions, that of classification, must deal with the lack of invariance of speech and involves assigning time-varying stretches of input onto some mental categories serving as intermediate representations in the lexical mapping process. The other function, which copes with the quasi-continuous nature of speech and the absence of a unique and reliable marker of the word boundaries, is called segmentation. It plays a role in finding the correct alignment between the intended words in the signal and lexical representations. In this regard, Norris and Cutler (1985) insisted that segmentation is logically distinct from classification: whereas classification entails necessarily some form of segmentation, segmentation does not require classification.

In view of this distinction, the original claim that the syllable constitutes a "perceptual unit" is ambiguous, and indeed, more recent theorizing has envisaged the syllable either as a classification unit or as a segmentation point. For instance, Cutler and Norris (1988) proposed that in English, the onsets of strong syllables (i.e. syllables that contain an unreduced vowel) provide alignment points for the lexical mapping process, without the signal being categorized

into syllabic units.

In contrast, for French (as well as the other Romance languages), the dominant conception is that the syllable constitutes the basic classification unit in speech perception and lexical access (Mehler, 1981; Mehler, Dupoux, & Seguí, 1990; Seguí, 1984; Seguí, Dupoux, & Mehler, 1990). In particular, Mehler et al. (1990) argued that the signal is recoded and categorized prelexically in terms of syllable-sized units. In their view, "speech is segmented into elementary units that roughly correspond to the syllable. […] Syllabic frames are recognized by a bank of syllabic analyzers. […]" (p. 255). In what follows, we briefly review the relevant evidence about the perceptual role of syllables in French and discuss it in the light of the segmentation/classification contrast introduced above.

An influential finding taken to favor prelexical syllabic classification comes from a seminal study by Mehler, Dommergues, Frauenfelder, and Seguí (1981) using the sequence detection task (see Frauenfelder & Kearns, 1996, for further details). In this study, French listeners were faster at sequence detection for targets that matched the first syllable (e.g. BA in BALANCE or BAL in BALCON) than for targets which corresponded either to more or less than the first syllable (BA in BALCON or BAL in BALANCE). This cross-over interaction between target type (CV vs. CVC) and word structure (CV. words vs. CVC. words) has been called the *syllable effect*.

In follow-up experiments (Cutler, Mehler, Norris, & Seguí, 1986), in which English and French listeners detected sequences in both French and English materials, the language-specific nature of the syllable effect was demonstrated. However, the difference lay less in the acoustic/phonetic realization of the words in the two languages than in the perceptual mechanisms developed in the two groups of listeners. In fact, the syllable effect was shown by French participants whatever the language of input, whereas no such effect occurred in either language for English participants. The sequence detection paradigm has since been used in

numerous cross-linguistic investigations (see Cutler, 1993, for an overview) which have contributed to the wide acceptance of the idea that the syllable is a unit of classification in French (see Kolinsky, 1998, for detailed review and discussion).

Yet, several arguments suggest that the sequence detection studies do not provide conclusive evidence in favor of prelexical syllabic classification. First, as pointed out by Morais and Kolinsky (1992; see also Kolinsky, 1998), the pattern of response latencies obtained by Mehler et al. (1981) does not exactly fit with the predictions of a syllabic classification account. Indeed, under the sequentiality assumption implied by this explanation, one would expect the detection of CV targets in CVC words (BA in BALCON) to yield shorter latencies than the detection of CVC targets in CV words (BAL in BALANCE), because listeners need to identify the second syllable in the latter case but not in the former.

Second, the sequence detection task requires listeners to explicitly segment the carrier and evaluate the match with the target (e.g. Frauenfelder & Kearns, 1996). Hence, the observed effects may not reflect early perceptual processing, but conscious segmentation strategies applying at the post-recognition level (Kolinsky, 1998; Morais, 1985; Morais, Content, Cary, Mehler, & Seguí, 1989). In addition, the finding that the occurrence of the effect is restricted to slow participants (cf. Dupoux, 1993; Sebastián-Gallés, Dupoux, Seguí, & Mehler, 1992) casts doubt on the mandatory status of a perceptual syllable coding process.

Third, recent data indicate that the syllable effect may be linked to specific acoustic-phonetic properties of some types of consonants, a result which considerably reduces the theoretical relevance of the syllable effect. In Mehler et al.'s (1981) experiment, all words started with a plosive, followed by the same vowel (/a/), and either /l/ or /r/ as pivotal consonant. Content, Meunier, Kearns, and Frauenfelder (in press) had French speakers detect CV or CVC sequences at the beginning of bisyllabic pseudowords varying in syllable structure and pivotal consonant. Overall, the study failed to replicate the syllable effect. A clear

crossover interaction was however observed for liquid pivotal consonants under target-blocking conditions, hence suggesting that the syllable effect may be restricted to the use of liquid consonants. In addition, regression analyses showed that phoneme duration accounted for a large proportion of the variance for CVC target detection, suggesting that participants were reacting rather directly to phonemic throughput.

The sensitivity of native-speakers of Romance languages to syllabic structure has been demonstrated using other paradigms. For instance, using the attentional allocation technique introduced by Pitt and Samuel (1990), Pallier, Sebastián-Gallés, Felguera, Christophe, and Mehler (1993) examined the effect of a manipulation of listeners' expectations. They showed that phoneme monitoring was facilitated when participants were induced to expect targets in a fixed position within the syllabic structure (i.e. either at the coda as in TAC<u>T</u>ILE or at the onset as in TA<u>B</u>LEAU), regardless of their absolute phonemic position (for example in the "onset" case, CA<u>P</u>RICE as well as PRO<u>B</u>LÈME; see also Tabossi, Collina, Mazetti, & Zoppello, 2000, for similar results in Italian).

Illusory conjunctions of dichotic stimulus fragments have also been tested for syllable effects. Using CVCV stimuli, Kolinsky, Morais, and Cluytens (1995) examined the probability of illusory word percepts (e.g. BIJOU, *jewel*, or KOTON, *cotton*) resulting from the erroneous recombination of various parts of the two inputs. They observed that syllable migrations (<u>BI</u>TON/KO<u>JOU</u>) were more frequent than migrations involving two adjacent but not tautosyllabic fragments (KI<u>J</u>ON/<u>B</u>OT<u>OU</u>).

Finally, in the study mentioned above, Tabossi et al. (2000, Experiment 5) used the semantic cross-modal priming paradigm. Three-phoneme fragments, excised from trisyllabic base words, were presented as primes in a visual lexical decision task. Targets were related or unrelated to the words from which the priming excerpts were extracted. In support of the syllabic access hypothesis, a significant (facilitatory) priming effect was found only when the

syllabic structure of the prime matched that of the base word related to the target (e.g. /si.l/, from SILENZIO, *silence,* vs. /sil./, from SILVESTRE, *silvan*, for the target RUMORE, *noise*; /sɔl./, from SOLDATO, *soldier*, vs. /sɔ.l/, from SOLARE, *solar*, for the target GUERRA, *war*).

All these results indicate that listeners do exploit some properties of the syllabic structure of speech. However, none of them constitutes direct evidence in favor of the hypothesis that the syllable serves as an access code to the mental lexicon. While Tabossi et al.'s (2000) cross-modal priming results are certainly consistent with a syllabic view of lexical access, they remain open to alternative explanations, as the authors themselves noted.

Furthermore, recent data obtained using syllabification tasks challenge the view that French listeners have clear and concordant intuitions about syllable boundaries. Content, Kearns, and Frauenfelder (in press) showed that French listeners are not consistent in their syllabification of simple CVCV words. Participants had to repeat either the first or the second part of the same CVCV stimulus words. A clear dissociation was observed. The vast majority of second-part responses included the intervocalic consonant (e.g. BALLON > LON), as predicted by all phonological analyses of French. But first-part responses split nearly evenly between CV and CVC responses (BALLON > BA or BAL). Moreover, the tendency to produce closed (CVC) first-part responses was more manifest for more sonorous consonants, and also was influenced by orthographic gemination (e.g. BALLON vs. PALAIS). Similar findings were obtained with pre-literate and literate children (Content, Dumay, & Frauenfelder, 1999).

The observation that second-syllable responses in syllabification tasks nearly always began with the consonant suggests that syllable onsets constitute reliable segmentation points in the signal and led us to the view that, rather than serving as classification units, syllables are primarily involved in segmentation. More precisely, syllable onsets help the listener segment the speech flow. Given that syllable onsets and word onsets frequently coincide, a syllable onset segmentation strategy may also be a useful heuristic for determining potential word

beginnings. Hence, the detection of syllable onsets would help lexical segmentation by providing the lexical search mechanisms with privileged alignment points in the signal (or its prelexical representation). We therefore proposed the Syllable Onset Segmentation Heuristic (SOSH; Content et al., in press; Frauenfelder & Content, 1999; see also Content, Dumay, & Frauenfelder, 2000), according to which listeners take any syllable onset to be a possible word onset, and use it as an alignment point for the lexical mapping process. In this framework, the goal of segmentation is less to permit speech classification than to find likely alignments between the (presumably infra-syllabic) classification units and the lexical entries in order to recover the intended parse.

The present study tested the SOSH hypothesis in a direct way by assessing the role of syllable onsets and offsets in spoken word recognition. We used the word-spotting task (Cutler & Norris, 1988) in which participants are required to discover monosyllabic words embedded in longer nonsense carriers. The word-spotting task has proven very useful for studying the role of various segmentation cues for word recognition in continuous speech (see McQueen, 1996, for further details). Indeed, it has been shown to be sensitive to variables like metrical structure and stress (Banel & Bacri, 1994; Cutler & Norris, 1988; Vroomen & de Gelder, 1997a), lexical competition (McQueen, Norris, & Cutler, 1994; Norris, McQueen, & Cutler, 1995; see also Dumay, Frauenfelder, & Content, 2000), phonotactics (Gaygen & Luce, submitted; McQueen, 1998; van der Lugt, in press; Weber, 2000), and allophonic variations (Dumay, Content, & Frauenfelder, 1999; Kirk, 2000; Yerkey & Sawusch, 1993). Here, the syllabification of the carriers was manipulated by varying the nature of the medial consonant clusters, so that the syllable boundary could be aligned or misaligned with one boundary of the embedded target word. If the speech stream is classified into syllabic units as assumed by Mehler and co-workers (e.g. Mehler et al., 1990), similar costs should be obtained for onset and offset misalignment, since in both cases the target word would be misaligned with a

boundary between two classification units. In contrast, if, as proposed by SOSH, speech is classified into smaller size units, and syllable onsets constitute preferred alignment points, only onset alignment should be crucial for target recognition, and consequently, misalignment effects should be stronger for finally- than initially-embedded targets.

Experiment 1 was aimed at assessing the effect of onset misalignment, and used finally-embedded target words, whose onset was either aligned (e.g. such as LAC, *lake*, in ZUN.LAC) or misaligned (ZU.GLAC) with the syllable boundary. Experiment 2 examined the effect of offset misalignment, with initially-embedded targets, either aligned (such as JUP(E), *skirt*, in JUP.TÈCHE) or misaligned (JU.PLOUNE) with the syllable boundary. Finally, Experiment 3 was a replication of Experiment 2, but using a different carrier set for the aligned condition (JUP.CHONTE). Whereas Experiment 2 used medial consonant clusters that exist in French at a syllable boundary but within words, Experiment 3 employed clusters that occur only across word boundaries.

Experiment 1

Method

Participants. Thirty-two students (mean age: 27, range: 19-41) from the University of Geneva participated in the experiment, as volunteers or as part of an introductory course in psycholinguistics. They were all native speakers of French and reported no hearing or speech disorder.

Materials. Thirty French content words of CVC structure, starting with a liquid (e.g. LAC, /lak/, *lake*) and with a mean frequency of 7,361 (range: 34-27,436) occurrences per 100 millions, were selected using Brulex, a lexical database for French (Content, Mousty, & Radeau, 1990). These words appeared as targets embedded in final position in CVCCVC bisyllabic nonce carriers.

The main variable of interest was the syllabification of the carriers, which was manipulated by changing the first consonant of the medial cluster. In one condition, the two consonants were separated by a syllable boundary (i.e. CVC.CVC), so that the target word was aligned with the onset of the second syllable of the carrier (e.g. ZUN.LAC). In the other condition, the two consonants were both assigned to the onset of the second syllable (i.e. CV.CCVC), so that the target was misaligned (by one phoneme) with the onset of the second syllable (e.g. ZU.GLAC).

Because existing descriptions of syllabification disagree for many classes of consonant clusters, we compared several proposals and selected clusters for which at least four out of five were in agreement (see Goslin & Frauenfelder, in press, for further details). The only clusters unanimously considered as non-separable consisted of a subset of the OBstruent-LIquid clusters. This subset, referred to as the OBLI class by Dell (1995), includes the combinations of a stop or the fricative /f/ or /v/ with the liquid /ʀ/ or /l/, excluding /tl/ and /dl/ clusters, illegal at word onset. In the misaligned condition, the OBLI clusters consisted of /dʀ/, /fʀ/, /kʀ/, /pʀ/, /tʀ/ and /vʀ/ for the 21 targets beginning with /ʀ/, and /bl/, /fl/, /kl/ and /gl/ for those beginning with /l/. In the aligned condition, the consonant clusters were /lʀ/, /mʀ/ and /nʀ/ for targets starting with /ʀ/, and /ml/, /nl/ and /ʀl/ for those starting with /l/. Except /nl/ and /nʀ/, all the C.C clusters were attested within a French morpheme.

The carriers were all phonotactically legal in French. To eliminate the influence of lexical cues on segmentation, none of the initial CV, CVC and CVCC portions constituted words. Similarly, to avoid lexical competition effects, the initial CVCC portion was compatible with no longer word, and the final CCVC portion never constituted a word or the beginning of a longer word.

Thirty nonce CVCCVC bisyllabic strings in which no CVC word was embedded were created as filler trials along the same principles as those followed for the carriers. For half of the

fillers, the syllable boundary fell within the medial consonant cluster (e.g. GUR.LINQUE); for the other half, it fell just before (e.g. NA.DRIC). To avoid any association between the identity of the consonant sequence and the presence of a target, the distribution of the consonant clusters in the fillers was roughly identical to those in the carriers. Finally, 20 additional trials similar to the experimental materials and in which no experimental target appeared served as practice block.

All the carriers –listed in Appendix 1– and fillers were pronounced in a sound-proof booth by a male native speaker of French, unaware of the goals of the experiment. The speaker was required to read each string silently before saying it aloud. The stimuli, recorded via a Sennheiser ME-80 microphone, were directly stored onto a Macintosh II CI computer at a 16,000 Hz sampling rate with 16-bit analog-to-digital conversion, using the Sound Designer II software. Each item was then transferred to a single sound file, and the duration of the carrier, target and liquid was measured from the waveform. On average, neither the carriers nor the targets differed in duration significantly between the aligned and the misaligned condition (682 ms (range: 539-775) vs. 674 ms (524-755), $t(29) = .57$, for carriers; 388 ms (287-475) vs. 384 ms (293-464), $t(29) = .53$, for the targets). In contrast, the liquid was found to be significantly longer in the CVC.CVC than in the CV.CCVC carriers (84 ms (55-106) vs. 71 ms (20-105), $t(29) = 3.56$, $p < .002$).

The carriers and fillers were assigned to two lists. Each included 30 carriers and 30 fillers in a fixed pseudo-random order, with the constraint that no more than three carriers or three fillers occurred consecutively, and that two carriers in the same alignment condition never appeared in direct succession. Targets and fillers occurred in the same position across the two lists, whereas the alignment condition for a given target was varied from one list to the other. Alignment was counterbalanced over the lists such that each list contained 15 targets in each

condition. Each list was split up into two blocks of 30 trials, in which the items were distributed in nearly equal proportions.

Procedure. Stimulus presentation, timing and data collection were controlled by PsyScope 1.1 (Cohen, MacWhinney, Flatt, & Provost, 1993). The stimuli were presented with D/A conversion directly from the disk, and played to the participants at a comfortable level through Beyerdynamic DT-100 headphones. Participants were tested individually in a sound-proof booth. They were told that they would hear a list of nonsense strings, and that their task was to attempt to spot the real words embedded at the end of some of the strings, provided that these words began with one consonant. Participants were instructed to press a button (with their preferred forefinger) as soon as they spotted a word, and to repeat that word aloud immediately. Each trial started with a 100 ms warning signal which terminated 1000 ms before the stimulus. The time interval between the onset of two successive trials was 4.5 s. Reaction times corresponding to the button push were measured from carrier onset, and the oral responses were checked on-line by the experimenter for accuracy.

Sixteen participants were assigned to each list. The order of block presentation was balanced so that eight participants were presented with block 1 and then block 2, while this order was reversed for the eight other participants. Before the experiment started, two runs of the practice trials presented in two different orders were performed. The experiment lasted about 20 minutes.

Results and Discussion

Reaction times (RTs) and Error rates were used as dependent variables. Chronometrical analyses were based upon RTs with correct oral responses only. RTs were adjusted by subtracting carrier duration so as to measure from target offset. Error rates (7% overall) included omissions as well as target-bearing trials for which participants pressed the button but reported nothing or a wrong word. The rate of false detections, i.e., positive responses to

filler trials was 2.7%. Mean RTs (by participants) and mean error rates are reported in Table 1.

Table 1. Mean reaction times (RTs) measured from target offset and error rates for Experiment 1.

		Carrier Type	RTs (ms)	Ers (%)
Target	Aligned	zun.lac	361 (117)	3.3 (5.4)
	Misaligned	zu.glac	454 (153)	10.6 (9.1)
	Difference		93 $^{++}$	7.3 $^{++}$

Note. $^{++}$ indicates a difference statistically significant by participants and items.

On average, target words were spotted 93 ms faster when their onset was aligned with the onset of the second syllable of the carrier than when it was misaligned with it. Repeated measures analyses of variance (ANOVAs) performed on mean RTs by participants (F_1) and by items (F_2) with alignment as factor revealed that the alignment effect was highly significant ($F_1(1,31) = 15.45$, $MSE = 9{,}075.9$, $p < .001$; $F_2(1,29) = 7.97$, $MSE = 33{,}794.5$, $p < .01$). As regards accuracy, error rates were on average more than three times lower in the aligned than in the misaligned condition. ANOVAs similar to those performed on the RTs indicated that the alignment effect on the error rates was also significant ($F_1(1,31) = 16.32$, $MSE = 52.1$, $p < .001$; $F_2(1,29) = 5.38$, $MSE = 148.3$, $p < .03$). In sum, spotting a word embedded finally is both faster and easier when the onset of the word is aligned with the onset of a syllable than when this is not the case.

In the aligned condition, one third of the carriers contained a cluster which was only attested at a morpheme boundary (e.g. /nl/ or /nʀ/), while the rest included clusters which could be found within morphemes. In order to assess whether the alignment effect was primarily due to syllabic or morphological cues, post-hoc analyses were run to compare the

two subsets. A 116 ms alignment effect was obtained for the morphological clusters (334 vs. 450 ms, respectively for the aligned and misaligned conditions), and a 81 ms effect was observed for the syllabic clusters (370 vs. 451 ms). There was no interaction between alignment and type of cluster ($F_1 < 1$; $F_2(1,28) = 1.38$, N.S.). Similarly for error rates, the alignment by type of cluster interaction failed to reach significance ($F_1(1,31) = 3.73$, $MSE = 164.1$, $p = .063$; $F_2(1,28) = 1.77$, $MSE = 144.5$, $p < .2$), in spite of a larger difference for morphological clusters (3.1 vs. 16.3%, respectively, for the aligned and misaligned condition) than for syllabic clusters (3.4 vs. 7.8%). Yet, the interpretation of the latter trend is less straightforward, since it appears to be due to an increase in errors in the misaligned condition, rather than to a decrease in the aligned condition which would be expected if morphological clusters constitute stronger segmentation cues.

The main finding of this experiment is that the misalignment between word onset and syllable onset produces a strong and highly reliable processing cost on both RTs and error rates. It is worth noting that, due to design constraints, all targets started with a liquid (either /l/ or /ʀ/). Participants could have taken advantage of such a regularity and deployed ad hoc attentional strategies for locating the onset of the target word, which might have in turn washed out the effect.

In sum, the present results provide strong evidence for the use of the syllabic structure in lexical segmentation of speech. However, taken individually, they are compatible both with the syllabic classification hypothesis (cf. Mehler et al., 1990) and with SOSH (cf. Content et al., in press). In order to disentangle these two hypotheses, Experiment 2 was aimed at assessing the effect of misalignment between word offset and syllable offset. If the syllable serves as classification unit, offset misalignment should entail a processing cost of similar magnitude to onset misalignment. In constrast, if only the onsets of syllables are important for

recognition of words in continuous speech, no or little alignment effect should be found in the spotting of initially-embedded words.

Experiment 2

Method

Participants. Thirty-two students (mean age: 21, range: 18-29) from the University of Geneva participated in this experiment, as volunteers or as part of an introductory course in psycholinguistics. They were all native speakers of French and reported no hearing or speech disorder. None of them had taken part in Experiment 1.

The main manipulation here concerned the alignment between the end of the target words and the syllable boundary of the carriers (e.g. JUP(E), /ʒyp/, *skirt,* in JUP.TÈCHE for aligned vs. JU.PLOUNE for misaligned). Selection of the clusters followed the same criteria as those of Experiment 1. The OBLI clusters used in the misaligned condition consisted of /bʀ/, /dʀ/, /fl/, /gʀ/, /kl/, /kʀ/, /pl/, /tʀ/, /vl/ and /vʀ/. In the aligned condition, the clusters used were /kt/, /ft/, /vn/, /pt/, /bd/, /gz/, /dv/, /fk/, /kp/, /tk/, /tp/ and /vz/. Among the latter, the first six exist within French morphemes, whereas the six others can be found only at a morphemic boundary within polymorphemic words.

Materials and Procedure. Thirty CVC monosyllabic French content words (mean frequency: 6,741, range: 314-43,985) were used as targets inserted in initial position in CVCCVC bisyllabic nonce strings.

To avoid lexical competition effects, the CVCC initial portion of the carriers was never a word, nor the beginning of a longer word. Moreover, to avoid the influence of lexical knowledge, neither the final CVC portions nor the final CCVC portions constituted words, and the latter were nearly never compatible with longer words. Other characteristics of the carriers

and fillers were similar to those of Experiment 1. The target-bearing materials are presented in Appendix 1.

Carrier duration showed no significant difference between the two conditions (696 ms (588-840) vs. 691 ms (563-792); $t(29) = .37$). In contrast, targets were on average significantly shorter in the aligned than in the misaligned condition (255 ms (160-351) vs. 280 ms (157-377); ($t(29) = 2.99$; $p < .01$). The procedure was identical to that of Experiment 1, except that the participants were required to spot initially-embedded words provided that these ended with one consonant.

Results and Discussion

Table 2. Mean reaction times (RTs) measured from target offset and error rates for Experiment 2.

		Carrier Type	RTs (ms)	Ers (%)
Target	Aligned	jup.tèche	748 (229)	15.6 (9.2)
	Misaligned	ju.ploune	812 (238)	15.6 (10.8)
	Difference		64 $^{+}$	0

Note. $^{+}$ indicates a difference statistically significant by participants only.

RTs were adjusted by subtracting the duration of the initial CVC portion so as to measure from target offset. Overall, the error rate (omissions and manual responses with no oral response or incorrect oral response) was 15.6%, whereas the rate of false detections was 6.5%. Mean RTs (by participants) and mean error rates are reported in Table 2.

On average, target words were spotted 64 ms faster when their offset was aligned with the offset of the first syllable of the carrier than when it was misaligned with it. Repeated measures ANOVAs with alignment as factor revealed that this difference was significant by

participants but not by items ($F_1(1,31) = 6.93$, $MSE = 9{,}287.4$, $p < .02$; $F_2(1,29) = 2.10$, $MSE = 25{,}511.9$, $p < .2$). As regards accuracy, error rates were identical across conditions, and no effect of alignment was found (both $Fs < 1$). As in Experiment 1, post-hoc analyses assessed whether the effect of alignment was dependent on the (morphological vs. syllabic) nature of the C.C clusters. In this experiment, 50% of the carriers pertained to each cluster type. On RTs, the alignment effect was 107 ms for the morphological clusters (704 vs. 811 ms, for the aligned and misaligned conditions) but only 23 ms for the syllabic clusters (782 vs. 805 ms). However, the interaction between alignment and type of cluster was merely marginally significant by participants and not significant by items ($F_1(1,31) = 3.00$, $MSE = 18{,}772.4$, $p = .093$; $F_2(1,28) = 1.74$, $MSE = 24{,}873.0$, $p < .2$). On error rates, there was no interaction between alignment and type of cluster ($Fs < 1$; 18.3 vs. 20.0% for the aligned and misaligned conditions in morphological clusters, and 12.9 vs. 11.3% for the same conditions in syllabic clusters).

The effect of final misalignment observed in the present experiment was thus less clear-cut, and differed in two ways from that of initial misalignment found in Experiment 1. First, the final misalignment effect was restricted to RTs, whereas the effect of initial misalignment was found on both RTs and errors. Second, the final misalignment effect found on RTs was not statistically reliable across items but was only due to some particular clusters. In contrast, its initial counterpart, significant in both analyses, was fully reliable.

In order to assess whether the trend towards an alignment effect found in Experiment 2 was caused by the presence of morphological consonant clusters, as suggested at least by the means obtained in the post-hoc analyses, a third experiment used exclusively word-boundary clusters. These clusters never occur within words in French but are attested across word boundaries. We reasoned that such sequences, compared to the OBLI clusters, would indicate an obligatory word offset and would represent the strongest possible syllable-offset cues.

Hence, if syllable-offset alignment is important for lexical activation and segmentation, we would expect Experiment 3 to reveal a greater effect than that found in Experiment 2. In contrast, if the trend observed for the morphological clusters is spurious, no effect should be obtained in Experiment 3.

Experiment 3

Method

Participants. Thirty-two students (mean age: 22, range: 18-27) from the University of Geneva participated in this experiment, as volunteers or as part of an introductory course in psycholinguistics. They were all native speakers of French and reported no hearing or speech disorder. None of them had taken part in Experiment 1 or 2.

Materials and Procedure. The only difference between this experiment and Experiment 2 was the nature of the clusters used in the aligned condition (see Appendix 1). Here, we selected aligned clusters (/bg/, /dg/, /fp/, /gv/, /kʃ/, /pʃ/, /tv/, /vb/ and /vd/) that never appeared within words in French but could be found at word boundaries. Thus, in the typical example, JUP.TÈCHE became JUP.CHONTE.

On average, as in Experiment 2, targets were significantly shorter in the aligned than in the misaligned condition (255 ms (146-347) vs. 280 ms (157-377); ($t(29) = 4.15$; $p < .001$). In addition, by contrast to Experiment 2, carriers were also significantly longer in the aligned than the misaligned condition (723 ms (582-833) vs. 691ms (563-792); $t(29) = 2.31$, $p < .03$). As the latter difference was in opposite direction to the local variations of the initial CVC portion, it could only be explained by (local) variations in the final CVC.

Except for carriers in the aligned condition and the corresponding fillers, modified to maintain a similar distribution of medial clusters, the rest of the materials and procedure were identical to Experiment 2.

Results and Discussion

Two items (i.e. MAUVE and RAIDE) missed by more than 50% of the participants were excluded from the analyses. Again, RTs were adjusted by subtracting the duration of the initial CVC portion so as to measure from target offset. Overall, the error rate (omissions and manual responses with no oral response or incorrect oral response) was 12.5%. The rate of false detections was 7.3%. Mean RTs (by participants) and mean error rates are reported in Table 3.

Table 3. Mean reaction times (RTs) measured from target offset and error rates for Experiment 3.

		Carrier Type	RTs (ms)	Ers (%)
Target	Aligned	jup.chonte	815 (190)	10.7 (8.7)
	Misaligned	ju.ploune	849 (221)	14.3 (7.7)
	Difference		34	3.6 $^{+}$

Note. $^{+}$ indicates a difference statistically significant by participants only.

On average, targets were spotted 34 ms faster in the aligned than in the misaligned condition, but the effect of alignment was not significant ($F_1(1,31) = 1.54$, $MSE = 12{,}037.7$, $p < .25$; $F_2(1,27) = 1.40$, $MSE = 28{,}798.5$, $p < .25$). On the error rates, participants tended to miss fewer targets in the aligned than in the misaligned condition (difference: 3.6%) but this trend reached significance by participants only ($F_1(1,31) = 4.59$, $MSE = 44.4$, $p < .05$; $F_2 < 1$).

In sum the use of stronger alignment cues did not lead to a larger effect of alignment. Instead, neither the RT difference nor the trend observed on error rates appeared statistically reliable.

General Discussion

In the present study, three word-spotting experiments evaluated the role of syllable structure in the recognition of words in continuous speech. In Experiment 1, the effect of misalignment between word onset and syllable onset was investigated (e.g. ZUN.LAC vs. ZU.GLAC). A strong and highly reliable processing cost due to the misalignment was observed both on RTs and on error rates. Experiment 2 assessed the effect of offset misalignment between words and syllables (JUP.TÈCHE vs. JU.PLOUNE), using materials analogous to those of Experiment 1. In contrast to Experiment 1, a smaller effect, only significant by participant, was obtained on RTs, and no effect was found on the error rates. Experiment 3 re-examined the offset misalignment case, using a stronger aligned condition, in which the clusters employed exclusively occurred at word boundaries (JUP.CHONTE vs. JU.PLOUNE). Compared to Experiment 2, the difference obtained on RTs was markedly reduced and the effect of alignment no longer significant; on error rates, a small advantage, significant only by participants, appeared for the aligned condition. Taken together, the results indicate much stronger onset than offset alignment effects, in line with the SOSH hypothesis.

Although RTs were measured from target offset in all experiments, they were about 400 ms faster in Experiment 1 (final embedding) than in the two other experiments (initial embedding). The finding of a target position effect is not specific to our study (Banel & Bacri, 1995; McQueen, 1998; McQueen et al., 1994; Norris, McQueen, Cutler, & Butterfield 1997; van der Lugt, in press). Assuming that target identification takes about the same time in both embedding positions, the RT difference suggests that for initial embedding, other processes take place after listeners have identified the target. Hence, one might worry that a potential misalignment effect on RTs in the initial embedding experiments is obscured by later processing.

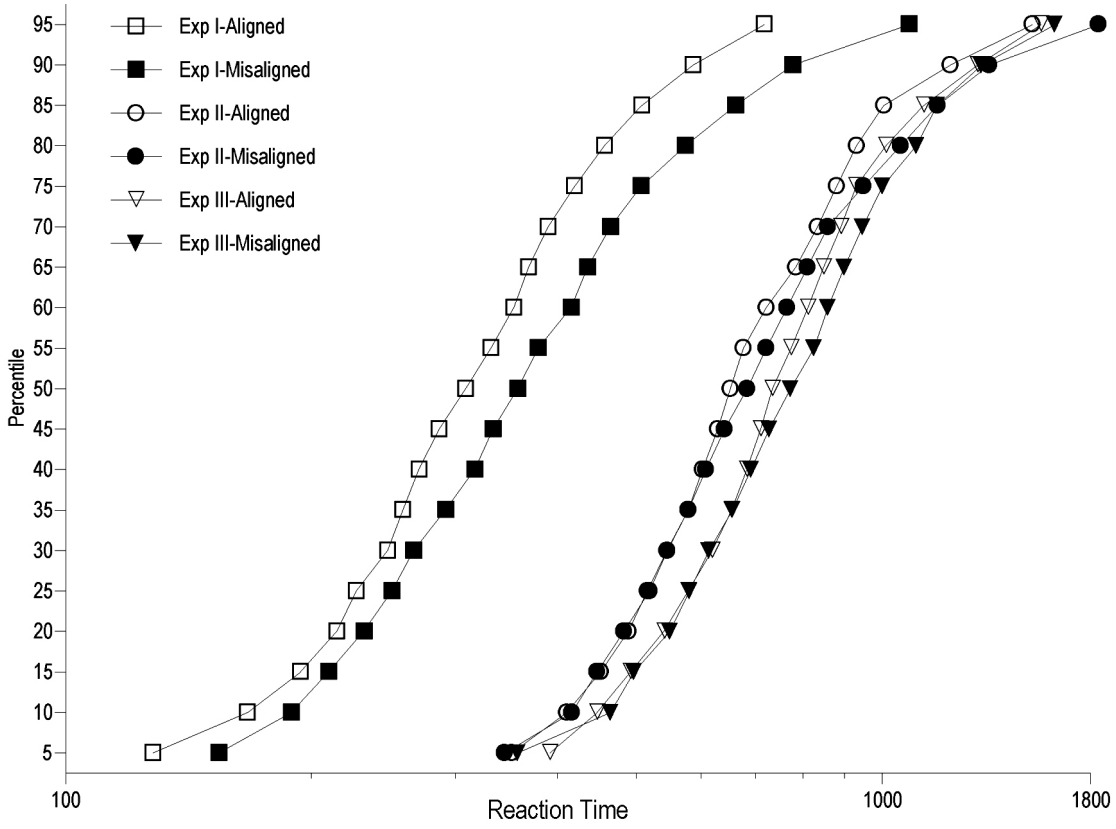

Figure 1. Cumulative spotting latency distributions for the aligned and misaligned targets for Experiments 1-3. The distributions were obtained by computing the 19 evenly spaced percentiles (from the 5th to the 95th), estimated from the raw RTs in each condition.

In order to evaluate this hypothesis, we checked whether misalignment effects appear for the fastest RTs and disappear for the longer latencies. We examined the distribution of raw RTs in each experiment, separately for the aligned and misaligned condition, and computed every fifth percentile. Figure 1 displays the cumulative latency distributions for the three experiments. Whereas the onset misalignment effect observed in Experiment 1 clearly extends over the whole range of the distribution, the offset misalignment effects only appear for the slower half of the distribution in Experiments 2 and 3. Hence, the distribution analysis provides absolutely no support to the view that the absence of a clear offset misalignment effect is due to late-occurring processes[1].

Another methodological issue stems from RT measurements. One could argue that measuring RTs from the offset (rather than from the onset) of the targets may have decreased the probability of finding a final misalignment effect, as for initial-embedding, targets were longer in the misaligned than in the aligned condition. Although this reasoning is correct, it does not explain the onset/offset asymmetry found on the error rates. Furthermore, measuring latencies from target offset, the standard practice in word-spotting experiments, provides an estimation of processing time from a point at which all the information required for target identification is available. This is not the case for latencies measured from target onset, as these are inevitably confounded with target duration (see McQueen, 1996).

Interestingly, other studies carried out in French provide little evidence in favor of an offset alignment effect. The only research that reported a significant effect (of 58 ms; Banel & Bacri, 1997) was aimed at assessing the relation between phonotactic legality and metrical cues to lexical segmentation. The authors had listeners spot initially-embedded CVC words in CVCCVC or CVCVC carriers (e.g. LAMP.ZOK vs. LAM.POK), bearing either an iambic (short-long), trochaic (long-short) or spondaic (long-long) rhythmic pattern. They observed a misalignment effect which did not interact with the manipulation of the metrical pattern. However, to provide the strongest possible cues to word segmentation, Banel and Bacri (1997) intentionally used very unusual or even "illegal" aligned clusters, in the sense that most violated the voicing agreement rule. Furthermore, it is likely that the misalignment effect was also enhanced by a lexical competition effect, since the target-overlapping final portion of the carriers was compatible with a larger set of cohort candidates in the misaligned (.POK) than in the aligned condition (P.ZOK). Hence, their finding of an offset misalignment effect is definitely not incompatible with the present data. Indeed, in four experiments quite similar to Experiments 2 and 3, Crouzet (2000) found either no effect or small effects of offset misalignment, similar in magnitude to those reported in the present study. When present, the

effect was never fully reliable, except in one experiment in which alignment was blocked, a situation that may well induce ad hoc strategic/attentional biases. In sum, there is little clear evidence that word/syllable offset misalignment introduces robust processing costs.

While the preferred interpretation of the onset/offset asymmetry is based on the assumption that only syllable onsets play a direct role in lexical segmentation, three alternative interpretations need to be discussed. First, as noted in the method sections of Experiments 2 and 3, initially-embedded targets were significantly shorter in the aligned than in the misaligned condition. Although such a durational difference is expected, since it reflects the propensity of French to keep syllable duration equivalent across various levels of structural complexity (e.g. Bartkova, 1988; Vaissière, 1977), it implies that aligned targets provide less processing time than misaligned targets, and it raises the possibility that the aligned targets contain less detailed acoustic-phonetic information than the misaligned targets. If this were the case, then potential final alignment effects could have been washed out due to a greater difficulty to identify targets in the aligned condition. To assess the plausibility of such an explanation, correlations between the magnitude of the alignment effect and the difference in target duration within each pair of carriers were calculated for Experiments 2 and 3. Contrary to the above explanation which predicts a negative correlation between the alignment effect and the durational difference, only weak and nonsignificant correlations were obtained (Experiment 2: $r(30) = -.18$, $p < .35$, for RTs; $r(30) = .11$, $p < .60$, for error rates; Experiment 3: $r(28) = .21$, $p < .30$, for RTs; $r(28) = .22$, $p < .30$, for error rates). Moreover, to make sure that a similar drop in the amount of acoustic-phonetic information was not responsible for reducing the alignment effect in Experiment 3 compared to Experiment 2, we examined whether the targets aligned with a morpheme-boundary in Experiment 2 were not shorter in Experiment 3. Recall that, as suggested by the post-hoc analyses, the alignment effect obtained in Experiment 2 was caused primarily by these targets. Contrary to the explanation

in terms of amount of acoustic information, the targets in question did not differ in duration across experiments ($t(14) = 1.69$, N.S.; 248 ms for Experiment 2, vs. 262 ms for Experiment 3). There is thus no evidence in favor of the idea that potential final alignment effects could have been compensated due to a greater difficulty to identify the targets in the aligned condition. This implies that the tendency observed in Experiment 2 of morpheme-boundary clusters to yield a larger RT alignment effect than their syllabic counterparts was likely due to random variations.

Secondly, in most of the previous studies as well as in the present one, the manipulation of onset vs. offset misalignment was confounded with the metrical structure. All carriers were stressed on the final syllable, producing the iambic pattern characteristic for bisyllables in French (e.g. Wenk & Wioland, 1982). This is confirmed in the present experiments by the difference in duration between initially- and finally-embedded targets ($t(58) = 9.96$; $p < .001$ for Experiment 2 vs. 1, and $t(58) = 8.93$; $p < .001$ for Experiment 3 vs. 1). Fortunately, the absence of a rhythm by phonotactics interaction reported by Banel and Bacri (1997) suggests that this confound cannot account for the onset/offset asymmetry.

Thirdly, several results obtained across various paradigms demonstrate that spoken word recognition involves inhibitory competition between overlapping lexical candidates (Dumay et al., 2000; Luce, Pisoni, & Goldinger, 1990; McQueen et al., 1994; Norris et al., 1995; Vroomen & de Gelder, 1995). Although our materials were carefully devised to avoid, as far as possible, contaminating effects of lexical competition, it was impossible to completely equate the aligned and misaligned conditions. For example, in Experiment 1, the nature of the manipulation entails that there are lexical candidates compatible with the .CCV portion (e.g., .GLA), in the misaligned condition, whereas none exists for the C.CV portions (e.g., N.LA) in the aligned condition. However, since in Experiment 1 the CCVC final portion of the carriers never constituted a word nor matched any longer word, it seems unlikely that the short-lived

activation of these candidates would be sufficient to explain the large onset alignment effect that we observed. Indeed, other data (Frauenfelder, Scholten, & Content, in press; Marslen-Wilson, 1993) indicate that activated lexical candidates incur rapid deactivation through bottom-up inhibition caused by a mismatching phoneme.

The present results are consistent with others obtained in Dutch and English. In a lexical decision task, Vroomen and de Gelder (1997b) assessed the activation of monosyllabic words embedded in longer items, by measuring the amount of cross-modal semantic priming visual words elicited, whether related or not to the embedded ones. They demonstrated that words aligned with the onset of strong syllables were activated (e.g. BOOS, angry, in FRAMBOOS, *raspberry*) but found no evidence of activation for words that were misaligned with the onsets of monosyllables (WIJN, *wine*, in ZWIJN, *swine*). In a study similar to the present one, McQueen (1998) showed that misalignment between word and syllable onsets (e.g. ROK, *skirt*, in FI.DROK vs. in FIM.ROK) produced a stronger processing cost than misalignment between word and syllable offsets (e.g. VEL, *skin*, in VELM.BRUL vs. in VEL.BRUL). As in the French study by Banel and Bacri (1997) cited above, McQueen (1998) also found that these phonotactic effects were not modulated by the metrical pattern, here induced by the presence of a full or a reduced vowel in the non-target portion of the carriers. Finally, the onset misalignment effect was recently replicated in English (e.g. PUN.LUCK vs. MAR.FLUCK), both with English monolinguals and German-English bilinguals (Weber, 2000).

In sum, the results of the present set of experiments, together with those reviewed above, indicate larger and more systematic alignment effects at syllable onset than offset. This asymmetry seems hard to reconcile with the hypothesis of prelexical syllabic classification (cf. Mehler et al., 1990), which would have predicted similar alignment effects at syllable onset and syllable offset. In contrast, it fits well with the hypothesis that syllables are used in speech recognition primarily as segmentation cues. Previous findings based on syllabification

tasks, outlined in the introduction, already suggested a predominant role for syllable onsets (Content et al., in press). The present study confirms and extends these findings using an on-line word recognition task. By providing direct evidence that syllable onsets are more important than syllable offsets for lexical alignment, it brings further support to the SOSH hypothesis, according to which syllable onsets constitute privileged reference points for segmentation and access to the lexicon. As Cutler and co-workers nicely conclude in a recent discussion, "the syllable appears to be the measuring stick against which viable and unviable parses of continuous speech are judged" (Cutler, McQueen, Norris, & Somejuan, 2001).

One counter-intuitive aspect of any view ascribing a functional role to syllables in spoken word recognition is that word and syllable boundaries do not always coincide in continuous speech. Since many words are multisyllabic, some syllable onsets are not word onsets. According to any syllabic segmentation strategy, such false alarms might be particularly damaging for the speech processor if the syllable boundary is the onset of another existing word (as MESSE, *mass*, in PROMESSE, *promise*, or BONE in TROMBONE), since the latter, being activated, could act as a competitor of the longer word. In agreement with the hypothesis of syllable-based segmentation, cross-modal priming (Isel & Bacri, 1999; Luce & Cluff, 1998; Shillcock, 1990) and perceptual identification findings (Cluff & Luce, 1990) demonstrate that the embedded item is indeed activated and more so if it spans an entire syllable (cf. Vroomen & de Gelder, 1997b). However, whether the complementary processing cost for the recognition of the multisyllabic embedding word does also occur is much less clear. In their lexical decision and shadowing experiments on bisyllabic words, Luce & Lyons (1999) found no effect of the lexical status of the second syllable (e.g. CHLORIDE vs. CHLORINE). One interpretation of these results is thus that the lexical processor does evaluate the parse involving the finally-embedded word because it is syllabically-aligned, but favors the longer word (cf. Swinney, 1981).

Another difficulty relates to resyllabification phenomena applying across word boundaries. In French, three types of phonological processes are assumed to entail resyllabification: elision, where two consonants are chained due to the deletion of an intervening schwa (as in "pas d' rôle", /pa.d'ʀol/, *no role*; "'" refers to the lost schwa); enchaînement, where a word-final consonant attaches to a following word-initial vowel or, in some cases, consonant (as in "sac anglais", /sa.kɑ̃.glɛ/, *British bag*); and liaison enchaînée, where a word-final latent consonant surfaces and is linked to a following word-initial vowel (as in "petit ami", /pə.ti.ta.mi/, *little friend*). In each of these cases, if syllabification rules apply to the whole phrases, word and syllable boundaries would not coincide, and an incorrect lexical alignment would be made on the basis of syllable onsets. However, various findings indicate that the pivotal consonant involved in elision (Fougeron & Steriade, 1997; Rialland, 1986), enchaînement (Dumay et al., 1999, 2000) or liaison (Dejean de la Bâtie, 1993; Spinelli, McQueen, & Cutler, 2000) is phonetically and perceptually different from its word-initial variant. For instance, in Dumay et al. (1999), we found that enchaînement sequences involving an OBLI cluster ("magique roche", /…ik#ʀɔʃ/, *magic rock*; "#" indicates the word boundary) differ in the duration of both the pre-boundary vowel and the liquid from sequences in which the cluster is lexically assigned to the second word ("demi-croche", /…i#kʀɔʃ/, *eighth note*). Furthermore, using the shared portions of these sequences (e.g. ik#ʀɔʃ vs. i#kʀɔʃ) as carriers in a word-spotting task on the CVC items, we demonstrated that listeners exploit these durational variations (or phonetic correlates of them) in on-line lexical segmentation. Thus, potential costs of word/syllable misalignments in continuous speech processing might be strongly alleviated if the segmentation heuristics is sensitive to fine-grained phonetic details. Moreover, it should be noted that syllable-based segmentation strategies such as SOSH (Content et al., in press) or that proposed by Cutler et al. (2001) are

not deterministic rules but heuristics, so that their effect could be modulated or compensated by other cues, such as lexical information. In line with this view, Fougeron, Frauenfelder, and Content (1999) compared lexical effects on phoneme monitoring, and found that the decrease in the lexical effect due to the insertion of an initial epenthetic consonant (e.g. "gacrobat") could be compensated when the inserted consonant had a lexical status (e.g. "l'acrobat"; see also Spinelli et al., 2000, for related findings).

One important implication of this research concerns the similarities and differences in processing across languages. Cutler and Norris (1988) argued that the metrical structure provides important constraints on word recognition in stress-timed languages such as English and Dutch. According to their proposal, the Metrical Segmentation Strategy (MSS), listeners would initiate lexical searches at the onsets of those syllables containing a full vowel, i.e. the strong syllables, whereas no lexical look-up would be triggered by weak syllables, which by definition contain a reduced vowel and never bear stress. Though it is implicit in the MSS hypothesis, segmenting the speech stream at strong syllables thus necessarily entails some syllable onset segmentation. Based on this view, and given that similar onset/offset alignment asymmetries were reported for French (this study) and Dutch (McQueen, 1998), one may speculate that the same onset alignment heuristic operates in the stress-timed languages and French, even though the way the rhythm of these languages constrains the lexical segmentation and mapping processes may greatly vary. If this appeared to be true, the Syllable Onset Segmentation Heuristic would constitute a general component of spoken word recognition, independent of the listener's language.

Acknowledgements

The work reported herein was supported by grants from the Direction Générale de la Recherche Scientifique – Communauté française de Belgique (A.R.C. 96/01-203), and from the Swiss National Fund for Scientific Research (Projects 1113-049698.96 and 1114-059532.99). It was carried out while the first author was research assistant of the National Fund for Scientific Research (Belgium). We thank David Ott for his voice, Marie-Hélène Banel for her help in testing participants, and Anne Cutler, Gareth Gaskell, James McQueen, Christophe Pallier, and two anonymous reviewers for fruitful comments or discussions.

Reprint requests should be sent to Nicolas Dumay, Psycholinguistics Research Group, Department of Psychology, University of York, Heslington, York YO10 5DD, United Kingdom. Email: n.dumay@psych.york.ac.uk

Footnotes

[1] A side-effect of this overall RT difference is that the variance was much larger in Experiments 2 and 3 than in Experiment 1, leading to a potential loss of statistical power in the former experiments. However, this potential problem only applies to the RT analysis but not to error rates.

References

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

Banel, M-H., & Bacri, N. (1995). Do metrical and phonotactic segmentation cues cooperate in spoken word recognition? *Proceedings of the 13th International Congress of Phonetic Sciences* (pp. 608-611), Stockholm, Sweden.

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

Bartkova, K. (1988). On the use of segmental duration in speaker-independent speech recognition systems. *Proceedings of Speech '88, 7th FASE Symposium* (pp. 763-770), Edinburgh, Scotland.

Cluff, M.S., & Luce, P.A. (1990). Similarity neighborhoods of spoken bisyllabic words. *Journal of Experimental Psychology: Human Perception and Performance*, **16,** 551-563.

Cohen J.D., MacWhinney B., Flatt M., & Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, **25,** 257-271.

Content, A., Dumay, N., & Frauenfelder, U.H. (1999). Segmentation syllabique chez l'enfant. *Actes des IIèmes Journées d'Etudes Linguistiques* (pp. 69-74), Nantes, France.

Content, A., Dumay, N., & Frauenfelder, U.H. (2000). The role of syllable structure in lexical segmentation in French: Helping listeners avoid mondegreens. In A. Cutler, J.M. McQueen, & R. Zondervan (Eds.), *Spoken Word Access Processes* (pp. 39-42), Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Content, A., Kearns, R.K., & Frauenfelder, U.H. (in press). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language*.

Content, A., Meunier, C., Kearns, R.K., & Frauenfelder, U.H. (in press). Sequence detection

in pseudowords: Where is the syllable effect? *Language and Cognitive Processes*.

Content, A., Mousty, P., & Radeau, M. (1990). Brulex: une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique*, **90,** 551-566.

Crouzet, O. (2000). *Segmentation de la parole en mots et régularités phonotactiques.* Unpublished doctoral dissertation. Université Paris V- René Descartes, France.

Cutler, A. (1993). Language-specific processing: Does evidence converge? In G.T.M. Altmann, & R. Shillcock (Eds.), *Cognitive models of speech processing: Second Sperlonga Meeting* (pp. 115-123). 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.

Cutler, A., McQueen, J.M., Norris, D., & Somejuan, A. (2001). The roll of the silly ball. In E. Dupoux et al. (Eds.), Language, brain and cognitive development: Essays in honor of Jacques Mehler (pp. 181-194). Cambridge, MA: MIT Press.

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

Dejean de la Bâtie, B. (1993). *Word boundary ambiguity in spoken French.* Unpublished doctoral dissertation, Monash University, Victoria, Australia.

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

Dumay, N., Content, A., & Frauenfelder, U.H. (1999). Acoustic-phonetic cues to word boundary location: Evidence from word-spotting. *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 281-284), San Francisco, CA.

Dumay, N., Frauenfelder, U.H., & Content, A. (2000). Acoustic-phonetic cues and lexical competition in segmentation of continuous speech. In A. Cutler, J.M. McQueen, & R. Zondervan (Eds.), *Spoken Word Access Processes* (pp. 127-130), Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Dupoux, E. (1993). The time course of prelexical processing: The syllable hypothesis revisited. In G.T.M. Altmann, & R. Shillcock (Eds.), *Cognitive models of speech processing: Second Sperlonga Meeting* (pp. 81-114). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Fougeron, C., & Steriade, D. (1997) Does deletion of French schwa lead to neutralization of lexical distinctions? *Proceedings of the 5th European Conference on Speech Communication and Technology*, Vol. 7, (pp. 943-937), University of Patras, Wire Communications Laboratory, Rhodes, Greece.

Fougeron, C., Frauenfelder, U.H., & Content., C. (1999) Is there an acrobat in nacrobat: How to juggle with boundary misalignment. *Proceedings of the XIth Conference of the European Society for Cognitive Psychology* (p. 148). Ghent: ESCOP/Academia Press.

Frauenfelder, U.H., & Content, A. (1999). The role of the syllable in spoken word recognition: Access or segmentation? *Actes des IIèmes Journées d'Etudes Linguistiques* (pp. 1-8), Nantes, France.

Frauenfelder, U.H., & Kearns, R.K. (1996). Sequence monitoring. In F. Grosjean & U.H. Frauenfelder (Eds.), *Spoken word recognition paradigms. A special issue of Language and Cognitive Processes* (Vol. **11,** pp. 665-673).

Frauenfelder, U.H., Scholten, M., & Content, A. (2001). Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes*.

Gaygen, D.E. & Luce, P.A., (2001). Troughs and bursts: Probabilistic phonotactics and lexical activation in the segmentation of spoken words in fluent speech. *Manuscript submitted for publication.*

Goslin, J., & Frauenfelder, U.H. (2000). A comparison of theoretical and human syllabification. *Language and Speech*, **44,** 409-436.

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

Kirk, C. (2000). Syllabic cues to word segmentation. In A. Cutler, J.M. McQueen, & R. Zondervan (Eds.), *Spoken Word Access Processes* (pp. 131-134), Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Kolinsky, R. (1998). Spoken word recognition: A stage-processing approach to language differences. *European Journal of Cognitive Psychology*, **10,** 1-40.

Kolinsky, R., Morais, J., & Cluytens, M. (1995). Intermediate representations in spoken word recognition: evidence from word illusion. *Journal of Memory and Language*, **34,** 19-40.

Luce, P.A., & Cluff, M.S. (1998). Delayed commitment in spoken word recognition: Evidence from cross-modal priming. *Perception and Psychophysics*, **60,** 484-490.

Luce, P.A., & Lyons, E.A. (1999). Processing lexically embedded spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, **25,** 174-183.

Luce, P.A., Pisoni, D.B., & Goldinger, S.D. (1990). Similarity neighborhoods of spoken words. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 122-147). Cambridge, MA: MIT Press.

Marslen-Wilson, W.D. (1993). Issues of process and representation in lexical access. In G.T.M. Altmann, & R. Shillcock (Eds.), *Cognitive models of speech processing: Second Sperlonga Meeting* (pp. 187-210). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

McQueen, J.M. (1996). Word-spotting. In F. Grosjean & U.H. Frauenfelder (Eds.), *Spoken word recognition paradigms.* A special issue of *Language and Cognitive Processes* (Vol. **11,** pp. 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. (1981). The role of syllables in speech processing: infant and adult data. *Philosophical Transactions of the Royal Society*, Series B, **295,** 333-352.

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

Mehler, J., Dupoux, E., & Seguí, J. (1990). Constraining models of lexical access: The onset of word recognition. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 236-262). Cambridge, MA: MIT Press.

Morais, J. (1985). Literacy and awareness of the units of speech: implications for research on the units of perception. *Linguistics*, **23,** 707-721.

Morais, J., & Kolinsky, R. (1992). In search of a paradigm for the study of speech perceptual codes. In *I Jornadas de Estudo dos Processos Cognitivos da Sociedade Portuguesa de Psicologica* (pp. 113-135). Lisboa: Astoria.

Morais, J., Content, A., Cary, L., Mehler, J., & Seguí, J. (1989). Syllabic segmentation and literacy. *Language and Cognitive Processes*, **4,** 57-67.

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

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., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, **34**, 191-243.

Pallier, C., Sebastián-Gallés, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language*, **32,** 373-389.

Pitt, M.A., & Samuel, A.G. (1990). Attentional allocation during speech perception: How fine is the focus? *Journal of Memory and Language*, **29,** 611-632.

Rialland, A. (1986). Schwa et syllabes en français. In L. Wetzels & E. Sezer (Eds.), *Studies in compensatory lengthening* (pp. 186-226). Dordrecht: Foris.

Savin, H.B., & Bever, T.G. (1970). The non-perceptual reality of the phoneme. *Journal of Verbal learning and Verbal Behavior*, **9,** 295-302.

Sebastián-Gallés, N., Dupoux, E., Seguí, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language*, **31,** 18-32.

Seguí, J. (1984). The syllable: A basic perceptual unit in speech processing. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and performance X: Control of Language processes* (pp. 165-181). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Seguí, J., Dupoux, E., & Mehler, J. (1990). The role of the syllable in speech segmentation, phoneme identification, and lexical access. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 263-280). Cambridge, MA: MIT Press.

Shillcock, R. (1990). Lexical hypotheses in continuous speech. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 24-49). Cambridge, MA: MIT Press.

Spinelli, E., McQueen, J.M., & Cutler, A. (2000). Lexical and acoustical factors in the resolution of liaison in French. *Abstracts of the 41st Annual Meeting of the Psychonomic Society* (Vol. 5, p. 66), New-Orleans, LO.

Swinney, D. (1981). Lexical processing during sentence comprehension: Effects of higher order constraints and implications for representation. In T. Myers, J. Laver, J. Anderson (Eds.), *The cognitive representation of speech* (pp. 201-209). Amsterdam: North-Holland.

Tabossi, P., Collina, S., Mazetti, M., & Zoppello, M. (2000). Syllables in the processing of spoken Italian. *Journal of Experimental Psychology: Human Perception and*

Performance, **26,** 758-775.

Vaissière, J. (1977). Premiers essais d'utilisation de la durée pour la segmentation en mots dans un système de reconnaissance. *Actes des VIIIèmes Journées d'Etudes sur la Parole* (pp. 345-352), Aix-en-Provence, France.

van der Lugt, A.H. (in press). The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics*.

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. (1997a). Trochaic rhythm in speech segmentation. *Abstracts of the 38th Annual Meeting of the Psychonomic Society* (Vol. 2, p. 73), Philadelphia, PA.

Vroomen, J., & de Gelder, B. (1997b). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **23,** 710-720.

Weber, A. (2000). The role of phonotactics in the segmentation of native and non-native continuous speech. In A. Cutler, J.M. McQueen, & R. Zondervan (Eds.), *Spoken Word Access Processes* (pp. 143-146), Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Wenk, B.J., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, **10,** 193-216.

Yerkey, P.N., & Sawusch, J.R. (1993). The influence of stress, vowel and juncture cues on segmentation. *Journal of the Acoustical Society of America*, **94,** 1880 (abstract).

Appendix 1: Target-bearing materials from Experiments 1, 2 and 3

Experiment 1: Final targets		
Aligned	Misaligned	*Target translation*
zun.lac	zu.glac	*lake*
gur.laine	zu.claine	*wool*
zin.lame	ti.glame	*sword*
zan.linge	gu.flinge	*linen*
gum.loupe	da.bloupe	*lens*
zam.lourd	ti.blourd	*heavy*
zir.luge	na.bluge	*sledge*
zèm.lune	cau.blune	*moon*
gueur.lutte	na.clutte	*fight*
zeum.race	gu.prace	*race*
zum.rage	zon.trage	*rage*
til.raide	na.draide	*stiff*
zeun.rame	co.prame	*oar*
zin.rampe	zé.vrampe	*ramp*
zul.rare	zé.trare	*rare*
zan.reine	da.creine	*queen*
gum.rêve	co.prêve	*dream*
nal.rhume	zi.frhume	*cold*
nam.riche	co.vriche	*rich*
zal.ride	quin.fride	*wrinkle*
nal.rire	gu.vrire	*smile*
gul.rive	gu.drive	*shore*
jan.robe	na.frobe	*dress*
zan.roche	zu.froche	*rock*
tim.role	zam.prôle	*role*
jal.rose	ti.drose	*rose*
zim.rouge	za.crouge	*red*
nan.route	da.froute	*road*
tin.ruche	da.pruche	*hive*
jam.rude	jo.trude	*rough*

Experiments 2 and 3: Initial targets			
Aligned 2	Aligned 3	Misaligned	*Target translation*
bec.tide	bec.chanfe	bè.clonche	*beak*
boeuf.quipe	bœuf.panque	boeu.flège	*beef*
bot.quère	bot.vasse	bo.truge	*boot*
chef.couge	chef.poute	chè.flupe	*boss*
chec.til	chec.chonle	chè.clide	*check*
chic.tose	chic.choule	chi.clonte	*fashionable*
choc.tus	choc.chide	cho.clume	*shock*
chut.pabe	chut.vèque	chu.trofe	*fall*
cod.vouche	cod.guile	co.drinche	*code*
coc.poeuf	coc.chimpe	co.cleute	*cock*
côt.quige	côt.vaf	cô.tranle	*coast*
dat.quize	dat.voule	da.trigue	*date*
dout.pigue	dout.vonque	dou.trule	*doubt*
guid.vate	guid.gace	gui.drouche	*guide*
jup.tèche	jup.chonte	ju.ploune	*skirt*
lamp.tuse	lamp.chog	lam.plosse	*lamp*
lang.jèpe	lang.vof	lan.gric	*tongue*
mauv.nasse	mauv.binque	mau.vrache	*mauve*
meut.poffe	meut.vimpe	meu.troufe	*pack*
nap.tine	nap.cholle	na.plite	*tablecloth*
pent.pache	pent.vuque	pen.treuse	*slope*
raid.vonce	raid.gonte	rai.drupe	*stiff*
rêv.zole	rêv.dace	rê.vluque	*dream*
rob.dac	rob.gane	ro.brine	*dress*
roc.tupe	roc.chune	ro.clinve	*rock*
rout.caphe	rout.vine	rou.trème	*road*
tic.tève	tic.chupe	ti.crouve	*twitch*
toc.tume	toc.chif	to.cluve	*hat*
typ.tinche	typ.chette	ty.plive	*type*
veuf.tal	veuf.pite	veu.flompe	*widowed*

Note. The items are given according to the orthographic rules of French. The written form of targets embedded initially is however not always respected.

Appendice 2.1: Porteurs de l'Expérience 1 (et stimuli des Expériences 2 et 3) et énoncés d'origine.

Structure	Obstruante	Liquide	Stimulus/ Porteur	Enoncé de type …C#C…	Enoncé de type …#CC…
VCCVC	voisée	/l/	aublate	quand l'aube latte	cadeau blatte
"	"	"	aublaisse	quand l'aube laisse	marteau blesse
"	"	"	aubloque	quand l'aube loque	bateau bloque
"	"	"	uglasse	subjugue lasse	menu glace
"	"	"	uglande	conjugue lande	bossu glande
"	"	"	uglève	subjugue lève	fichu glaive
"	"	"	uglobe	conjugue lobe	fichu globe
"	"	"	uglotte	subjugue lotte	tenu glotte
"	"	/ʀ/	aubrouille	quand l'aube rouille	canaux brouillent
"	"	"	aubroute	quand l'aube route	chevaux broutent
"	"	"	ondrame	féconde rame	maison drame
"	"	"	ondrôle	immonde rôle	chanson drôle
"	"	"	ugrape	subjugue râpe	venu grappe
"	"	"	ugreine	conjugue reine	reçu graine
"	"	"	ugrêve	subjugue rêve	voulu grève
"	"	"	ugrive	conjugue rive	reçu grive
"	sourde	/l/	aplaine	échappe laine	saga pleine
"	"	"	apleur	échappe leur	pacha pleure
"	"	"	iflac	nocif lac	demi-flaque
"	"	"	iflamme	récif lame	demi-flamme
"	"	"	iflèche	craintif lèche	demi-flèche
"	"	"	iflutte	rétif lutte	demi-flûte
"	"	"	ouplente	chaloupe lente	genou plante
"	"	"	ouplonge	chaloupe longe	jaloux plonge
"	"	/ʀ/	aprobe	échappe robe	moka probe
"	"	"	icrampe	dorique rampe	demi-crampe
"	"	"	icroche	magique roche	demi-croche
"	"	"	ifrire	natif rire	rôti frire
"	"	"	ifroc	actif roc	petit froc
"	"	"	oufriche	marlouf riche	bambou friche
"	"	"	ouprose	chaloupe rose	papou prose
"	"	"	utrace	minute race	venu trace
CVCCV	voisée	/l/	baglu	bague ludique	bas glucose
"	"	"	bomblan	bombe lancée	bon blanchi
"	"	"	fugla	fugue lacrymale	fut gladiateur
"	"	"	jambla	jambe lampiste	gens blanchis

Structure	Obstruante	Liquide	Stimulus/ Porteur	Enoncé de type …C#C…	Enoncé de type …#CC…
"	"	"	langlu	langue lubrique	lent glucose
"	"	"	tomblan	tombe lanterne	ton blanc-bec
"	"	"	tublon	tube longuet	tu blondis
"	"	"	vagla	vague lamelle	vas glaner
"	"	/ʀ/	cavrè	cave rêvassante	cas vraisemblant
"	"	"	guidru	guide rubrique	Guy Drucker
"	"	"	mauvrè	mauve rêvassante	maux vraisemblant
"	"	"	modra	mode racée	mot dragueur[28]
"	"	"	mondra	monde raté	mon drapeau
"	"	"	raidra	raide rateau	raie drapée
"	"	"	sudra	sud râpé	su draper
"	"	"	vidru	vide rumeur	vit Drucker
"	sourde	/l/	banclou	banque loufoque	banc clouté
"	"	"	jupla	jupe latine	jus platine
"	"	"	lamplon	lampe longuette	lent plombier
"	"	"	louplin	loupe limpide	loup plaintif
"	"	"	nuclan	nuque langoureuse	nu clandestin
"	"	"	pomplan	pompe lanterne	pont planqué
"	"	"	saclou	sac loufoque	sa clouterie
"	"	"	souplon	soupe lombarde	sous plombage
"	"	/ʀ/	botrin	botte ringarde	bot trinqueur
"	"	"	doutrin	doute ringard	doux trinqueur
"	"	"	fentri	fente ridée	faon tricheur
"	"	"	goutri	goutte rigide	goût tribal
"	"	"	patrou	patte rouquine	pas troublant
"	"	"	pentro	pente rocheuse	paon tropique
"	"	"	tantrou	tante roublarde	temps troublant
"	"	"	tètrou	tête rouquine	tes troupeaux

Note. Les stimuli sont retranscrits selon les règles de correspondance grapho-phonologique du français.

[28] Ici, la voyelle orthographique o en syllabe ouverte (e.g., "mot") était prononcée dans sa variante /ɔ/ de manière à éviter une trop grande disparité avec la voyelle en syllabe fermée (e.g., "mort").

Appendice 2.2: Remplisseurs de l'Expérience 1 et énoncés d'origine.

Structure	Obstruante	Liquide	Remplisseur	Enoncé de type …C#C…	Enoncé de type …#CC…
VCCVC	voisée	/l/	aublague	quand l'aube lague	oiseau blague ✓
"	"	"	aublouse	quand l'aube louse ✓	manteau blouse
"	"	"	aubluff	quand l'aube loeuf	moineau bluff ✓
"	"	"	uglaire	subjugue laire ✓	menu glaire
"	"	"	uglaise	conjugue laise	venu glaise ✓
"	"	"	uglauque	subjugue lauque ✓	paru glauque
"	"	"	uglêbe	conjugue lêbe	dodu glêbe ✓
"	"	"	uglose	conjugue lose	connu glose ✓
"	"	/ʀ/	aidrill	barmaid rile	lacet drill ✓
"	"	"	andrague	lavande ragues ✓	prenant dragues
"	"	"	aubribe	quand l'aube ribe	rateau bribe ✓
"	"	"	aubringue	quand l'aube ringue	marteau bringue ✓
"	"	"	ugreffe	conjugue reffe	Manu greffe ✓
"	"	"	ugrêge	subjugue rêge ✓	déçu grêge
"	"	"	ugrêle	subjugue rêle ✓	menu grêle
"	"	"	ugriffe	conjugue riffe	Manu griffe ✓
"	sourde	/l/	aplainte	étape linte ✓	état plainte
"	"	"	aplaire	échappe laire	coca plaire ✓
"	"	"	iflemme	chétif lemme ✓	petit flemme
"	"	"	ifleuve	canif leuve	joli fleuve ✓
"	"	"	iflic	natif lic ✓	petit flic
"	"	"	iflingue	actif lingue ✓	petit flingue
"	"	"	ouplage	chaloupe lage ✓	bisous plage
"	"	"	ouplus	soucoupe luce	jaloux plus ✓
"	"	/ʀ/	amprof	estampe roffe	étant prof ✓
"	"	"	antruc	méchante ruque ✓	méchant truc
"	"	"	apreuve	échappe roeuve ✓	donna preuve
"	"	"	icrêpe	barrique rêpe ✓	demi crêpe
"	"	"	icrête	chimique rête	semi-crête ✓
"	"	"	ifraise	calife raise ✓	garnit fraise
"	"	"	ifrêle	motif rêle	souris frêle ✓
"	"	"	ifringue	datif ringue ✓	poly-fringues
CVCCV	voisée	/l/	caglu	cague ludique ✓	cas glucose
"	"	"	coublan	coube lenteur	cou blanchi ✓
"	"	"	doblon	dobe longueur ✓	dos blondi
"	"	"	fingla	fingue lamelle	fin glamour ✓
"	"	"	guiblan	guibe lancée	gui blanchi ✓
"	"	"	maglu	magues lunaires	mat gluant ✓

Structure	Obstruante	Liquide	Remplisseur	Enoncé de type …C#C…	Enoncé de type …#CC…
"	"	"	mongla	mongue lacée	mont glaçon ✓
"	"	"	vublon	vube lombaire ✓	vu blondin
"	"	/ʀ/	bonvrè	bonve rêvassante ✓	bon vraisemblant
"	"	"	gadru	gade rugueuse ✓	gars Drucker
"	"	"	ganvrè	ganve rêvassante	gant vraisemblant ✓
"	"	"	jondra	jonde racornie ✓	jonc draconnien
"	"	"	noudrin	noude ringarde	nous dringuelle ✓
"	"	"	pidru	pide rutilante ✓	pi drupacé
"	"	"	toudra	toude râpée	tout drapé ✓
"	"	"	tudra	tude ragot ✓	tu draguais
"	sourde	/l/	bouplon	boupe longtemps	bout plombé ✓
"	"	"	fonclou	fonque loufoque ✓	fond clouté
"	"	"	fouplan	foupe lanterne	fou planté ✓
"	"	"	liplan	lipe lanterne ✓	lit planté
"	"	"	mouclan	mouque langoureuse ✓	mou clandestin
"	"	"	moupli	moupe lissée	mou plissé ✓
"	"	"	puclan	puc languissant	pu clandestin ✓
"	"	"	vimplon	vimpe lombaire	vin plombé ✓
"	"	/ʀ/	bintron	binte ronfleuse	bain tronqué ✓
"	"	"	dutri	dute ricanante ✓	du trituré
"	"	"	gaintron	gainte ronfleuse ✓	gain tronqué
"	"	"	lintrou	linte rouquine	lin trouvé ✓
"	"	"	moutrin	moute rincée ✓	mou trinqueur
"	"	"	nutro	nutte rococo ✓	nu tropical
"	"	"	rintrou	rinte rouquine ✓	rein troupeau
"	"	"	tautri	taute rituelle	taux tripoteur ✓

Note. Les items sont retranscrits selon les règles de correspondance grapho-phonologique du français. Le signe "✓" indique celui des deux énoncés qui a été choisi au hasard pour être utilisé.

Appendice 3: Distributions des TRs par condition pour les Expériences 1, 4, 7 et 8.

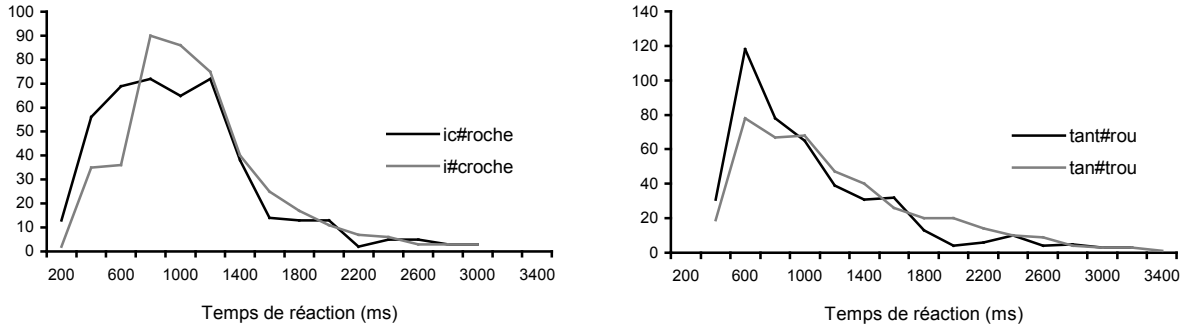

Figure 9. Distributions des TRs en fonction de l'alignement dans l'Expérience 1. A gauche: Cibles CVC enchâssées en finale. A droite: Cibles CVC enchâssées à l'initiale.

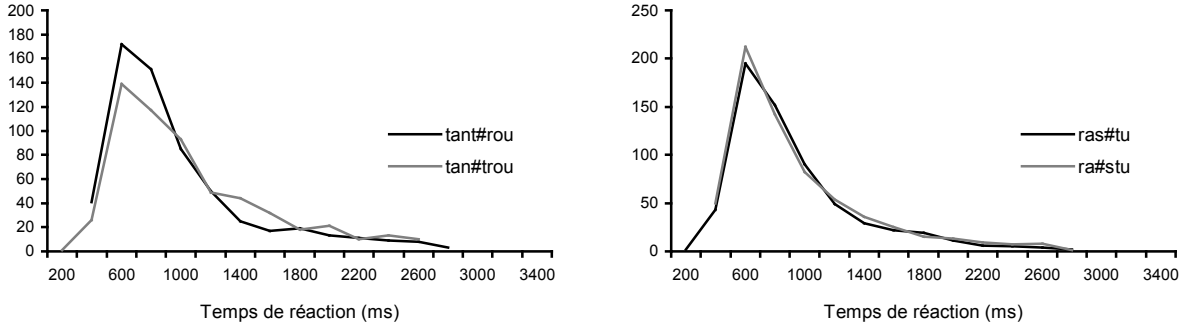

Figure 10. Distributions des TRs en fonction de l'alignement dans l'Expérience 4 (cibles CVC ou CCVC). A gauche: Porteurs à groupe consonantique OBLI. A droite: Porteurs à groupe consonantique SOB.

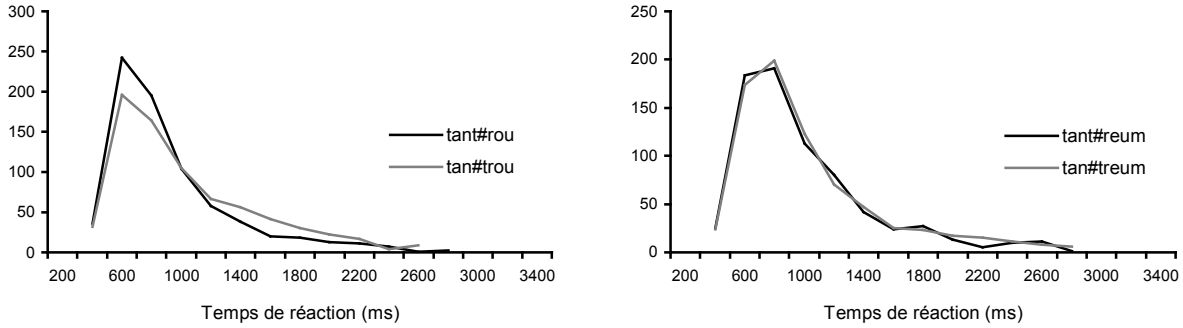

Figure 11. Distributions des TRs en fonction de l'alignement dans l'Expérience 7 (cibles CVC ou CCVC). A gauche: Porteurs à finale mot. A droite: Porteurs à finale pseudomot.

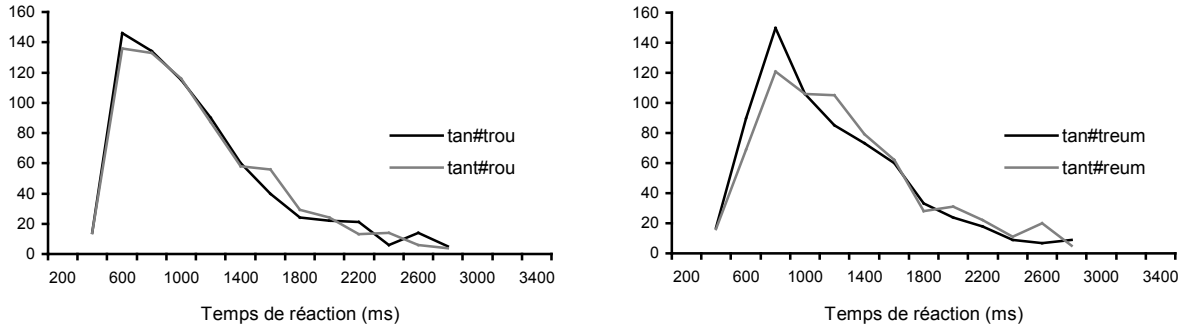

Figure 12. Distributions des TRs en fonction de l'alignement dans l'Expérience 8 (cibles CV ou CCV). A gauche: Porteurs à finale mot. A droite: Porteurs à finale pseudomot.

Appendice 4.1: Porteurs de l'Expérience 4 (et stimuli de l'Expérience 5) et énoncés d'origine.

Groupe consonantique	Structure	Porteur	Enoncé de type …C#C…	Enoncé de type …#CC…
SOB	CVCCV	baisphé	baisse phénoménale	baie sphéroïdale
"	"	bosca	bosse capricieuse	bot scabreux
"	"	chansco	chance colossale	chant scholastique
"	"	chasca	chasse cacophonique	chat scaphandrier
"	"	denspon	danse pontificale	dent spongieuse
"	"	fasté	face tétanisée	fa stéréo
"	"	fispon	fils pompier	fit spontanément
"	"	gospa	gosse pathétique	go spacieux
"	"	mastan	masse tempérante	mat standard
"	"	mespa	messe pathétique	mes spaghettis
"	"	minsca	mince cataracte	main scarifiée
"	"	mousté	mousse ténébreuse	mou sténographe
"	"	pinsphé	pince phénicienne	pain sphéroïdal
"	"	pousphé	pousse féconde	pou sphérique
"	"	rastu	race tutélaire	rat stupéfait
"	"	tasphé	tasse fêlée	tas sphérique
"	CCVCCV	brasque	brasse querelleuse	bras squelettique
"	"	brousti	brousse tibétaine	brou stimulant
"	"	cuispé	cuisse pétillante	cuit spécialement
"	"	fronsphé	fronce phénoménale	front sphéroïdal
"	"	glastu	glace tubulaire	glas stupéfiant
"	"	grasca	grasse catholique	gras scapulaire
"	"	gresphé	graisse fétide	grès sphérique
"	"	grosca	grosse carapace	gros scarabée
"	"	niesphé	nièce féministe	niais sphénisque
"	"	plascan	place cantonale	plat scandinave
"	"	pluspo	plus potentiel	plu sporadiquement
"	"	poisphé	poisse féconde	poids sphérique
"	"	prespé	presse périodique	prêt spéculatif
"	"	suista	Suisse talentueux	suis stagiaire
"	"	trespé	tresse pékinoise	très spécial
"	"	troustan	trousse temporaire	trou standard
OBLI	CVCCV	doutrin	doute ringard	doux trinqueur
"	"	bêtro	bête robuste	baie tropicale
"	"	langra	langue râpeuse	lent grappillage
"	"	lampré	lampe rétractable	lent préparatif
"	"	pentro	pente rocailleuse	paon tropical

Groupe consonantique	Structure	Porteur	Enoncé de type …C#C…	Enoncé de type …#CC…
"	"	guidru	guide rupestre	gui drupacé
"	"	jupro	jupe rococo	jus protidique
"	"	botrin	botte ringarde	bot trinqueur
"	"	bombri	bombe ricochante	bon bricoleur
"	"	cavrè	cave rêvassante	cas vraisemblable
"	"	jambra	jambe raccourcie	gens braconnant
"	"	modra	mode rapiéciée	mot draconnien
"	"	mondra	monde ratiboisé	mon drapeau
"	"	patrou	patte rouquine	pas troublant
"	"	sacro	sac rococo	sa croquignole
"	"	tantrou	tante roublarde	temps troublant
"	CCVCCV	craintron	crainte ronchonnante	crin trompeur
"	"	crêpro	crêpe rôtie	craie protéiforme
"	"	fricrin	fric ringard	fri craintivement
"	"	fritrou	frite rouquine	fri troupeau
"	"	frondra	fronde racommodée	front drapé
"	"	grèvrè	grève rêvassante	grès vraisemblable
"	"	grapro	grappe rôtie	gras protubérant
"	"	grècri	grec ridicule	grès criblé
"	"	grèfran	greffe remboursée	grès franciscain
"	"	grifran	griffe rentrée	gris framboise
"	"	plaintri	plainte ricanante	plein tripot
"	"	plantri	plante rituelle	plan trituré
"	"	placran	plaque rentable	plat cranté
"	"	trêvrè	trêve rêvassante	très vraisemblable
"	"	trompro	trompe robuste	tronc protubérant
"	"	troupré	troupe réticente	trou précancéreux

Note. Les séquences porteuses sont retranscrites selon les règles de correspondance grapho-phonologique du français.

Appendice 4.2: Remplisseurs de l'Expérience 4 et énoncés d'origine.

Groupe consonantique	Structure	Remplisseur	Enoncé de type …C#C…	Enoncé de type …#CC…
SOB	CVCCV	bansphé	bance féministe	banc sphérique ✓
"	"	binstu	bince tutélaire	bain stupéfait ✓
"	"	cansti	cance tibétaine ✓	camp stimulant
"	"	chousta	chouce talentueuse ✓	choux stagiaire
"	"	dinspon	dince pompeux ✓	dain spongieux
"	"	dusphé	duce féconde ✓	du sphénisque
"	"	finspé	fince périodique ✓	fin spéculative
"	"	fouscan	fouce cantonale ✓	fou scandinave
"	"	ganspo	gance potiquet	gant spoliateur ✓
"	"	lousphé	louce phénoménale	loup sphéroidal ✓
"	"	néstan	néce temporaire	nez standard ✓
"	"	rispa	rice pathétique	riz spacieux ✓
"	"	tansphé	tanse phénicienne ✓	temps sphérique
"	"	tinsque	tince querelleuse	thym squelettique ✓
"	"	tusca	tuce cataracte	tu scarifié ✓
"	"	vinsco	vince colossale ✓	vin scholastique
"	CCVCCV	blansque	blance quenouille	blanc skelettique ✓
"	"	bleusphé	bleuce féconde ✓	bleu sphérique
"	"	blonsté	blonce ténébreuse ✓	blond sténographe
"	"	clousca	clouce catholique ✓	clou scarabée
"	"	cransté	crance tétanisée ✓	cran sténographe
"	"	cruspon	cruce pompiste	cru spontané ✓
"	"	drasphé	drace fêlée ✓	drap sphéroidal
"	"	flanspo	flance potentielle	flanc sporadique ✓
"	"	flustan	fluce tempérante	flux standard ✓
"	"	frinsca	frince capricieuse ✓	frein scapulaire
"	"	grispé	grice pékinoise ✓	gris spécial
"	"	gruscan	gruce cantonale	grue scandinave ✓
"	"	plinspon	plince pontificale ✓	plein spongieux
"	"	plonsphé	plonce fétide	plomb sphénisque ✓
"	"	pristu	price tubulaire	prix stupéfiant ✓
"	"	trinsphé	trince phénicienne	train sphérique ✓
OBLI	CVCCV	bintron	binte ronfleuse ✓	bain tronqué
"	"	coubri	coube ricochante	coup bricolé ✓
"	"	dobri	dobe rituelle ✓	do bricolé
"	"	dougri	dougue ripaille	doux grippé ✓
"	"	dutri	dute ricanante	du trituré ✓

Groupe consonantique	Structure	Remplisseur	Enoncé de type …C#C…	Enoncé de type …#CC…
"	"	foncran	fonque rancunière ✓	fond cranté
"	"	foupré	foupe réparée ✓	fou préparé
"	"	jondra	jonde racornie ✓	jonc draconnien
"	"	lipré	lipe réticente	lit prétendu ✓
"	"	moutrin	moute rincée ✓	mou trinqueur
"	"	nutro	nutte rococo	nu tropical ✓
"	"	pidru	pide rutilante	pi drupacé ✓
"	"	rintrou	rinte rouquine	rein troupeau ✓
"	"	tautri	taute rituelle ✓	taux tripoteur
"	"	tavrè	tave rêvée	tas vraisemblant ✓
"	"	toudra	toude râpée ✓	tout drapé
"	CCVCCV	bleupro	bleupe rococo	bleu protidique ✓
"	"	bloncrin	blonque ringarde	blond craintif ✓
"	"	cloudra	cloude râpée ✓	clou drapé
"	"	crinpro	crimpe rotissante	crin protubérant ✓
"	"	critron	crite rompue ✓	cri trompeur
"	"	floutrin	floute ringarde ✓	flou trinqueur
"	"	flupri	flupe ricanante ✓	flux privilégié
"	"	froncri	fronque ripée ✓	front criblé
"	"	frutri	frute riquiqui	fru tricoté ✓
"	"	gluvrè	gluve raidissante ✓	glue vraisemblante
"	"	grintro	grinte rocaillleuse	grain tropical ✓
"	"	gropré	groppe récalcitrante ✓	gros prétentieux
"	"	plavrè	plave rêvassante	plat vraisemblable ✓
"	"	plifran	pliffe rentière ✓	pli franquiste
"	"	tronfran	tronffe rempante	tronc framboisé ✓
"	"	troucran	trouque remplie	trou cranté ✓

Note. Les items sont retranscrits selon les règles de correspondance grapho-phonologique du français. Le signe "✓" indique celui des deux énoncés qui a été utilisé après avoir été choisi au hasard

Appendice 5.1: Enoncés de deux mots et phrases porteuses utilisés dans l'Expérience 6.

Groupe consonantique	Séquence commune	PFL	Phrases porteuses
OB + /l/	baglu	C#C	Christian m'offrit une bague ludique en forme de clown.]
		#CC	Robert mangea un bas glucose aux fibres molles.]
"	banclou	C#C	Jo recensa vingt banques loufoques dans son carnet.
		#CC	Léo tomba d'un banc clouté en plein hiver.
"	bomblan	C#C	Le pilote évita la bombe lancée sans aucune difficulté.
		#CC	Chaque soldat recevra un bon blanchi par envoi recommandé.
"	fugla	C#C	Faire une fugue lacrymale consiste à partir en pleurant.]
		#CC	Artaud fut gladiateur dans le dernier film de Renoir.]
"	jamblan	C#C	Ils échangèrent une jambe lampiste contre un pied boiteux.
		#CC	Ils engagèrent des gens blanchis sans savoir qu'en faire.
"	jupla	C#C	Hélène portait une jupe latine taillée sur mesure.]
		#CC	Il en sortit un jus platine de couleur marron.]
"	lamplon	C#C	Un prisonnier déroba trois lampes longuettes avant de rentrer.
		#CC	Le préposé rameuta les lents plombiers d'un coup de sifflet.
"	langlu	C#C	Le caméléon déroula sa langue lubrique en un clin d'oeil.
		#CC	Le sportif affamé mangea son lent glucose en un clin d'oeil.
"	louplin	C#C	Le médecin prit une loupe limpide au verre très épais.]
		#CC	Mon papa vit un loup plaintif au poil très épais.]
"	nuclan	C#C	Le veinard caressait la nuque langoureuse de la belle Isabelle.]
		#CC	Ce Chinois revendait les nus clandestins de la belle Isabelle.]
"	pomplan	C#C	Ces mineurs utilisent une pompe-lanterne en plein jour.
		#CC	Les résistants montrèrent le pont planqué aux Anglais.
"	saclou	C#C	L'officier enleva tout sac loufoque sur ordre du colonel.
		#CC	Le menuisier revendit sa clouterie sur ordre du ministère.
"	souplon	C#C	L'enfant put déguster la soupe lombarde aux poivrons.]
		#CC	L'abri fut consolidé sous plombage intensif.]
"	tomblan	C#C	Il éguise des tombe-lanternes à chevillette.]
		#CC	Il s'agit de ton blanc-bec aux cheveux longs.]

Groupe consonantique	Séquence commune	PFL	Phrases porteuses
"	tublon	C#C	Ecoute ces six tubes longuets dans tes hauts-parleurs.
		#CC	Ecris-moi si tu blondis dans ton prochain pli.
"	vagla	C#C	Voici de vagues lamelles à trois trous.]
		#CC	Souris et vas glâner quelques pièces.]
OB + /ʀ/	botrin	C#C	Le cow-boy chaussa ses bottes ringardes d'une seule main.
		#CC	Il se fit à son pied-bot trinqueur sans problème.
"	cavrè	C#C	Vian dépeignit une cave rêvassante aux murs transpirants.]
		#CC	Nous avions parlé d'un cas vraisemblant de murs transparents.]
"	doutrin	C#C	Ce chef émettait des doutes ringards en son temps.
		#CC	Mon pére me semblait un doux trinqueur en son temps.
"	fentri	C#C	La biche se cacha dans une fente ridée de la roche.]
		#CC	Les enfants dessinaient un faon tricheur en savates.]
"	goutri	C#C	L'infirmière déposa une goutte rigide fluorescente.]
		#CC	Le curé exhibait un goût tribal très prononcé.]
"	guidru	C#C	Cette firme édita deux guides-rubriques l'an dernier.
		#CC	Le gendarme retrouva Guy Drucker sain et sauf.
"	mauvrè	C#C	Son pull était d'un mauve rêvassant tâché de jaune.]
		#CC	Laurent souffrait de maux vraisemblants déjà soignés.]
"	modra	C#C	Jeanne adore la mode racée du grand tailleur.]
		#CC	J'apprenais les mots dragueurs du grand savant.]
"	mondra	C#C	L'auteur décrit un monde raté sans pleurer.
		#CC	Les soldats hissaient mon drapeau dans la brume.
"	patrou	C#C	Le chat présenta sa patte rouquine pour une caresse.
		#CC	Le chef avançait d'un pas troublant vers le saloon.
"	pentro	C#C	L'imprudent dégringola d'une pente rocheuse très abrupte.]
		#CC	La police subtilisa un paon tropique très prisé.]
"	raidra	C#C	Le jardinier se prit un raide rateau en pleine figure.
		#CC	Le requin attaqua la raie drapée sans dire un mot.

Groupe consonantique	Séquence commune	PFL	Phrases porteuses
"	sudra	C#C	Nous marchions vers ce sud râpé d'un vert éclatant.]
		#CC	Notre couturière sut draper notre bel enfant.]
"	tantrou	C#C	Aurélie est une tante roublarde du Brabant.]
		#CC	Il nâquit en ces temps troublants d'avant guerre.]
"	tètrou	C#C	Le marin passa sa tête rouquine par le hublot.
		#CC	Le berger a gardé tes troupeaux toute la journée.
"	vidru	C#C	Il propage des vides rumeurs à chaque fois.
		#CC	Le président vit Drucker à chaque fois.
/s/ + OB	baisphé	C#C	Les cours subissent une baisse phénoménale depuis lundi.
		#CC	Cet arbre portait des baies sphéroïdales depuis avril.
"	bosca	C#C	Il lui poussa une bosse capricieuse dans le dos.
		#CC	Il avait un pied-bot scabreux étant gosse.
"	chansco	C#C	Il bénéficia d'une chance colossale de pendu.]
		#CC	Le coeur récita un chant scholastique d'Angleterre.]
"	chasca	C#C	Les renards fuirent cette chasse cacophonique infernale.]
		#CC	Le cirque exhibe un chat scaphandrier très habile.]
"	denspon	C#C	Le monde entier vit une danse pontificale improvisée.]
		#CC	Le dentiste extrait une dent spongieuse déracinée.]
"	fasté	C#C	Le coupable cachait sa face tétanisée de ses deux mains.
		#CC	Le guitariste lança un fa stéréo sans nous prévenir.
"	fispon	C#C	J'ai vu ton fils pompier sur les lieux.
		#CC	Ta soeur le fit spontanément sans savoir.
"	gospa	C#C	Ils cachèrent trois gosses pathétiques dans une grange.
		#CC	Je voudrais un go spacieux pour moi seul.
"	mastan	C#C	Il dut ajouté une masse tempérante importante.]
		#CC	L'équipe remonta le mat standard du navire.]
"	mespa	C#C	Le curé célébra une messe pathétique aujourd'hui.
		#CC	J'aurais bien aimé manger mes spaghettis aujourd'hui.

Groupe consonantique	Séquence commune	PFL	Phrases porteuses
"	minsca	C#C	J'ai souffert d'une mince cataracte comme un martyre.
		#CC	Il nettoie la main scarifiée à l'aide d'éther.
"	mousté	C#C	Nous nageons dans une mousse ténébreuse sans consistance.]
		#CC	Il a muté le mou sténographe du quatrième.]
"	pinsphé	C#C	Cette archéologue trouva des pinces phéniciennes en argent.]
		#CC	Cette boulangère fabriquait un pain sphéroïdal très prisé.]
"	pousphé	C#C	Nous avions repiqué des pousses fécondes dans le jardin.
		#CC	Le coiffeur attrapa le pou sphérique avec une pince.
"	rastu	C#C	Ils se proclamaient race tutélaire de la colonie.]
		#CC	Il dessina un rat stupéfait à la queue très courte.]
"	tasphé	C#C	La dame s'est débarassée des tasses fêlées à émail vert.]
		#CC	Le magicien construisit un tas sphérique fait de grains bruns.]

Note. Les séquences communes sont retranscrites selon les règles de correspondance grapho-phonologique du français. La position de la frontière lexicale (PFL) est indiquée par le signe "#". Le signe "]" indique les phrases dans lesquelles la partie en amont de l'énoncé s'accorde avec le premier terme de celui-ci; dans les phrases non-marquées, cette partie s'accorde avec le verbe.

Appendice 5.2: Enoncés de deux mots et phrases porteuses utilisés dans l'Expérience 6b.

Identité de la liquide dans les groupes OBLI	Phrases porteuses
/l/	Voici de vagues lamelles/balises à trois trous.]
"	Un prisonnier déroba trois lampes longuettes/tombales avant de rentrer.
"	Le veinard caressait la nuque langoureuse/pendulaire de la belle Isabelle.]
"	Le pilote évita la bombe lancée/dentée sans aucune difficulté.
"	Le médecin prit une loupe limpide/teintée au verre très épais.]
"	Le caméléon déroula sa langue lubrique/buveuse en un clin d'oeil.
"	L'officier enleva tout sac loufoque/touffu sur ordre du colonel.
"	L'enfant put déguster la soupe lombarde/tombale aux poivrons.]
"	Jo recensa vingt banques loufoques/touffues dans son carnet.
"	Il éguise des tombe-lanternes/dentelées à chevillette.]
"	Ils échangèrent une jambe lampiste/dansante contre un pied boiteux.
"	Hélène portait une jupe latine/taquine taillée sur mesure.]
"	Faire une fugue lacrymale/batailleuse consiste à partir en pleurant.]
"	Ecoute ces six tubes longuets/gondoles dans tes hauts-parleurs.
"	Christian m'offrit une bague ludique/durale en forme de clown.]
"	Ces mineurs utilisent une pompe-lanterne/camphrée en plein jour.
/ʀ/	Vian dépeignit une cave rêvassante/belliqueuse aux murs transpirants.]
"	Son pull était d'un mauve rêvassant/belliqueux tâché de jaune.]
"	Nous marchions vers ce sud râpé/gâché d'un vert éclatant.]
"	Le marin passa sa tête rouquine/coupante par le hublot.
"	Le jardinier se prit un raide rateau/bateau en pleine figure.
"	Le cow-boy chaussa ses bottes ringardes/pimbêches d'une seule main.
"	Le chat présenta sa patte rouquine/coupante pour une caresse.
"	La biche se cacha dans une fente ridée/pilée de la roche.]
"	L'infirmière déposa une goutte rigide/quinée fluorescente.]
"	L'imprudent dégringola d'une pente rocheuse/cochère très abrupte.]
"	L'auteur décrit un monde raté/battu sans pleurer.
"	Jeanne adore la mode racée/gaffeuse du grand tailleur.]

Identité de la liquide dans les groupes OBLI	Phrases porteuses
"	Il propage des vides rumeurs/buveurs à chaque fois.
"	Cette firme édita deux guides-rubriques/buvards l'an dernier.
"	Ce chef émettait des doutes ringards/pingres en son temps.
"	Aurélie est une tante roublarde/coudée du Brabant.]

Note. Le signe "]" indique les phrases dans lesquelles la partie en amont de l'énoncé s'accorde avec le premier terme de celui-ci; dans les phrases non-marquées, cette partie s'accorde avec le verbe.

Appendice 6.1: Enoncés dont sont extraits les porteurs des Expériences 7 et 8.

	Mot		Pseudomot	
Structure	C(C)VC#CV…	C(C)V#CCV…	C(C)VC#CVC…	C(C)V#CCVC…
CVCCV…	bague ludique	bas glucose	bague leultose	bas gleultose
"	banque loufoque	banc clouté	banque leuchpoque	banc cleuchpoque
"	bête rongeuse	baie trompeuse	bête reuznache	baie treuznache
"	bombe rivée	bon brimé	bombe reustouche	bon breustouche
"	cave rêvassante	cas vraisemblant	cave reuftante	cas vreuftant
"	chauve rêvassant	chaud vraisemblant	chauve reumlu	chaud vreumlu
"	dinde ravie	daim drapé	dinde reumliche	daim dreumliche
"	doute ringard	doux trinqueur	doute reufqueur	doux treufcoeur
"	fauve rival	faux vrillé	fauve reuspé	faux vreuspé
"	fente ridée	faon tricheur	fente reuftone	faon treuftone
"	fête roumaine	fait troublant	fête reumnale	fait treumnal
"	fugue latine	fût glacé	fugue leuchtine	fût gleuchtine
"	goutte rigide	goût tribal	goutte reumlique	goût treumlique
"	guide rubrique	Guy Drucker	guide reuspaque	gui dreuspac
"	jambe râpée	gens braqués	jambe reustive	gens breustifs
"	jupe ridée	jus prisé	jupe roeuznée	jus proeuzné
"	lac loufoque	la cloutier	lac leuftone	la cleuftone
"	lampe rétive	lent prélat	lampe roeuzbile	lent proeuzbile
"	ligue latine	lit glacial	ligue leuchtène	lit gleuchtène
"	loupe rigide	loup privé	loupe roeuzneuche	loup proeuzneuche
"	mode racée	mot drapé	mode reuftée	mot dreufté
"	monde raté	mont drapé	monde reustin	mont dreustin
"	nuque landaise	nu clenché	nuque leultaise	nu cleultais
"	patte rouquine	pas troublant	patte reugmine	pas treugmine
"	pente rocheuse	paon tropique	pente reugmate	paon treugmate
"	pompe ridée	pont privé	pompe roeuzbuche	pont proeuzbuche
"	raide rateau	raie drapé	raide reuftine	raie dreuftine
"	rive rêvassante	riz vraisemblant	rive reuzbaque	riz vreuzbac
"	route ronflante	roux tronqué	route reuftane	roux treuftane
"	rude rafale	ru drapé	rude reuftant	ru dreuftant
"	sac loufoque	ça clouté	sac leultoffe	ça cleultoffe
"	site routier	si trouvé	site reugmace	si treugmace
"	soupe ridée	sou prisé	soupe roeuzneule	sou proeuzneul
"	sud râpé	su draper	sud reunlé	su dreunler
"	tante roublarde	temps troublant	tante reumlite	temps treumlite
"	tête rouquine	tes troupeaux	tête reuftine	tes treuftines
"	tombe rivale	thon brisé	tombe reunliche	thon breunliche

	Mot		Pseudomot	
Structure	C(C)VC#CV…	C(C)V#CCV…	C(C)VC#CVC…	C(C)V#CCVC…
"	tube ringard	tu bringas	tube reuzbaque	tu breuzbaques
"	vague lacée	va glaner	vague leuftée	va gleufter
"	vide rumeur	vit Drucker	vide reusqueur	vit dreuscoeur
CCVCCV…	brave rigole	bras vrillé	brave reuftole	bras vreuftole
"	bringue lassante	brin glacé	bringue leuftace	brin gleuftace
"	drague lutine	drap glucose	drague leuchtine	drap gleuchtine
"	fringue lacée	frein glacé	fringue leuftée	frein gleufté
"	fronde racée	front drapé	fronde reumlée	front dreumlé
"	glande rugueuse	gland Drucker	glande reuscate	gland dreuscat
"	grave richesse	gras vrillage	grave reustale	gras vreustal
"	trêve rivale	trait vrillé	trêve reumnale	trait vreumnal
"	brique loupée	bris clouté	brique leuchtée	bris cleuchté
"	cloque loufoque	clos clouté	cloque leulpoque	clos cleulpoque
"	coiffe rendue	quoi franglais	coiffe reumlue	quoi freumlu
"	crainte rongeuse	crin trompeur	crainte reufnique	crin treufnique
"	crampe rivée	cran privé	crampe roeuznile	cran proeuznile
"	crêpe romane	crès profane	crêpe roeuznate	crès proeuznate
"	crique loufoque	cri clouté	crique leuchtoque	cri cleuchtoque
"	cuite rompue	cuit trompeur	cuite reumluse	cuit treumlu
"	flûte rivale	flux tribal	flûte reujnole	flux treujnole
"	frite rouquine	fri troublant	frite reusquine	fri treusquine
"	greffe rendue	grès français	greffe reumlue	grès freumlu
"	griffe rentrée	gris framboise	griffe reustiche	gris freustiche
"	grotte rivale	gros tricheur	grotte reuftache	gros treuftache
"	moite rougeur	mois troublant	moite reumlette	mois treumlette
"	plainte rivale	plein tripot	plainte reuftale	plein treuftal
"	plante rinçée	plan trinqueur	plante reumnée	plan treumné
"	plaque lancée	plat clenché	plaque leuchtée	plat cleuchté
"	pointe rigole	point tribal	pointe reustole	point treustol
"	soif rangée	soi frangin	soif reumlante	soi freumlant
"	suite roumaine	suis troublé	suite reuftée	suis treufté
"	tripe récente	tri présent	tripe roeuzneule	tri proeuzneul
"	trompe rangée	tronc prenchyste	trompe roeuznate	tronc proeuznate
"	trotte rigide	trot tribal	trotte reujnole	trot treujnole
"	troupe rétive	trou prédit	troupe roeuzneuche	trou proeuzneuche

Note. Les pseudomots sont retranscrits selon les règles de correspondance grapho-phonologique du français.

Appendice 6.2: Enoncés dont sont extraits les remplisseurs de l'Expérience 7.

	Mot		Pseudomot	
Structure	C(C)VC#CV…	C(C)V#CCV…	C(C)VC#CVC…	C(C)V#CCVC…
CVCCV…	binte ridée	cou brisé	chouque leuznaque	chant cleuftin
"	dobe raleuse	fin privée	fate reuchpoque	dent vreuspibe
"	dute ripée	gain tronqué	feupe roeuznarde	deux treufchouque
"	fonque loufoque	jeu tribal	linte reuchnante	gueux dreujnafe
"	foupe récente	jonc drapé	mique leulpode	nain treumlu
"	gade rugueuse	long troupeau	nibe reusperge	pain breustuque
"	moute rincée	main primaire	pide reufchouque	pou vreumlite
"	tave rêveuse	mât glucide	pougue leuchtase	rein dreustole
"	toude râpée	nom drapé	vingue leultome	saut dreumlute
"	vaugue lucide	taux vraisemblant	vute reuftaine	vous proeuzneuche
CCVCCV…	blanve rêveuse	bleu troublant	boigue leuftée	chien freunlote
"	cloupe rigide	blond clouté	brouque leulmine	coin proeuznif
"	flouque lancé	cru glacé	drimpe roeuznule	doigt proeuznoute
"	frute rivale	franc prélat	joite reumline	foin treujnice
"	glade rutilante	grain troqué	nièfe reuspante	fois treuspule
"	glufe rangée	pli prédit	noique leuchtine	loin vreuftose
"	plonte rivale	prompt glaneur	skide reuftale	pied treumlu
"	pneute romaine	train vrillé	vlante reustule	poids treunlet

Note. Les pseudomots sont retranscrits selon les règles de correspondance grapho-phonologique du français.

Appendice 6.3: Enoncés dont sont extraits les remplisseurs de l'Expérience 8.

Mot		Pseudomot	
C(C)V#CCV…	C(C)VC#CV…	C(C)V#CCVC…	C(C)VC#CVC…
boe français	blague latine	bla treuftite	bluff reuscan
co troublé	choc loupé	chet cleuftain	chef reuscain
cra brigand	chope rivale	cloe breujnole	chouette reustule
daux bravé	shoot rival	crou proeuzbèque	coq leultais
floe vraisemblant	code rugueux	dro gleuchpec	croute reuftèle
fraux drapé	dague lucide	fian treumlu	date reugmache
ja dragué	flaque loufoque	glet vreusté	diète reuftoche
jun trouvé	flingue labile	grou proeuzbuche	flic leulton
miais trivial	frappe rigide	na treuspique	globe reuftoule
na primaire	glotte roulante	proe vreustanche	phoque leufton
stè clenché	jatte rincée	pru dreumnot	quinte reuftole
stè prédit	scout rongé	sno breuspa	slave reujnu
tra précaire	spot ronflant	sta fleulcan	stade reuftan
tran trouveur	tif rincé	stan dreumnal	tic leulpot
tru clenché	top récent	tiè dreugmune	trac leuznac
wa tronqué	toque lancée	oua treumlu	trempe roeuznile
yo drapé	truffe landaise	yau treusponte	type roeuznache
zwa vraisemblant	veuve rêveuse	zi gleultais	zinc leuchtoux

Note. Les pseudomots sont retranscrits selon les règles de correspondance grapho-phonologique du français. A cause du nombre limité de mots CVC ne comportant pas de mot CV en français, la proportion de remplisseurs CVCCV(C), dans chaque condition, n'était pas toujours de 10/18, comme dans le matériel expérimental, mais respectivement de 5, 8, 8 et 10 sur 18.

Appendice 7: Test de fiabilité de la mesure du temps de réaction par le logiciel PsyScope (Cohen *et al.*, 1993).

Afin de quantifier, dans nos données chronométriques, la part de variabilité imputable au système de mesure lui-même, i.e. PsyScope 1.1 (dans Dumay *et al.*, 2002) ou 1.2.4 (dans les autres expériences), chaque dispositif expérimental a été soumis à un test de fiabilité. L'objet du présent appendice est de décrire ce test et de fournir une estimation de la précision avec laquelle nos temps de réaction ont été recueillis.

Description

La clé du test consiste à remplacer le participant par un système de feedback (Digital Delay Boss RPS10), qui garde en mémoire le stimulus reçu et le ré-émet en direction du boîtier de réponse après un intervalle de temps déterminé. Le stimulus est un bip de 20 ms. Le temps de réaction est mesuré à partir du début du stimulus jusqu'au moment où l'ordinateur reçoit la réponse (i.e. une copie du stimulus) générée par le système de feedback et transmise via le boîtier *ad hoc*. Si la mesure est exacte, la valeur observée du temps de réaction devrait donc être équivalente au temps, présumé constant, nécessaire pour parcourir la boucle [ordinateur – ampli – système de feedback – boîtier de réponse – ordinateur], additionné du délai introduit à l'aide du système de feedback.

Le test est basé sur 11 délais s'échelonnant de 0 à 800 ms, et sur les valeurs attendues correspondantes (sauf pour le délai 0, ce dernier servant à évaluer le temps minimal nécessaire au parcours de la boucle). Pour chaque valeur attendue, 20 temps de réaction ont été recueillis, et leur moyenne et écart-type ont été calculés.

Résultats

Dans la mesure où les trois dispositifs utilisés (respectivement dans Dumay *et al.* 2002, et dans les Expériences 1 et 4, et 3, 7 et 8) produisent des résultats analogues, nous n'en rapportons qu'un échantillon représentatif pris au hasard, en l'occurrence, les résultats obtenus avec le dispositif des Expériences 3, 7 et 8. Rappelons que ce dispositif était géré au moyen d'un Macintosh-G3 Power PC et de la version 1.2.4 du logiciel PsyScope, et que l'amplificateur était un Kenwood KAF-3010R.

Les résultats sont présentés au Tableau 13. Comme l'indique la condition dans laquelle aucun délai n'était introduit (première colonne), le temps minimal pour effectuer l'entièreté de la boucle était d'un peu moins de 19 ms. Quel que soit le délai utilisé (dans les autres colonnes), la moyenne des temps de réaction observés était donc très proche de la valeur attendue sur base de l'estimation du temps de boucle. Globalement, la différence était de .17 ms. En outre, et plus crucial encore, l'étendue de la dispersion des observations au tour de leur moyenne respective était tout à fait négligeable, l'écart-type le plus important étant seulement de .64 ms. Par ailleurs, comme en témoigne l'absence de corrélation entre l'écart-type et le temps de réaction moyen observé ($r = -.09$, n.s.), l'étendue de la dispersion n'était pas liée au niveau absolu du temps de réaction.

Conclusion

La part de variabilité imputable au système de mesure lui-même s'avère insignifiante. La précision et la fiabilité du logiciel PsyScope (Cohen *et al.*, 1993) sont donc particulièrement remarquables.

Tableau 13. En haut: Valeurs attendues du temps de réaction pour onze délais s'échelonnant de 0 à 800 ms. En bas: Temps de réactions observés, moyenne et écart-type (ms) sur 20 essais/valeur attendue, obtenus avec le dispositif des Expériences 3, 7 et 8.

Délai:	0	25	50	75	100	150	200	300	400	600	800	
T.R. attendu:	**?**	43.7	68.7	93.7	118.7	168.7	218.7	318.7	418.7	618.7	818.7	essai n°
T.R. observé:	19	44	70	94	119	168	219	319	418	619	819	1
	19	43	69	94	119	169	220	319	418	618	818	2
	18	44	69	94	119	169	219	318	419	618	819	3
	19	44	69	94	119	169	219	318	419	619	818	4
	19	43	69	94	119	170	220	319	419	618	818	5
	18	44	69	94	118	169	219	319	419	619	819	6
	19	45	69	93	120	170	219	318	419	618	819	7
	19	44	70	94	119	168	219	318	419	618	818	8
	18	44	69	94	119	169	219	319	418	619	818	9
	18	44	69	94	120	169	220	318	419	618	819	10
	19	44	69	93	119	169	219	318	419	618	819	11
	19	43	69	95	119	170	219	319	418	618	818	12
	18	44	69	93	118	170	220	319	419	619	818	13
	19	45	69	94	119	170	219	318	419	618	819	14
	19	44	70	94	119	169	219	318	418	619	818	15
	18	43	69	94	119	169	219	319	419	619	818	16
	18	45	69	94	118	169	219	319	419	619	819	17
	19	44	70	93	119	169	220	319	419	619	819	18
	19	43	69	94	119	169	218	318	418	619	818	19
	19	44	69	94	119	170	219	319	418	618	818	20
Moyenne:	**18.7**	43.9	69.2	93.9	119.0	169.2	219.2	318.6	418.7	618.5	818.5	
Ecart-type:	0.49	0.64	0.41	0.49	0.51	0.62	0.52	0.51	0.49	0.51	0.51	