

Chapitre 7

Niveau d'implémentation, Mode d'intégration

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

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

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

NIVEAU D'IMPLEMENTATION, MODE D'INTEGRATION

L'objet de ce travail était d'approfondir l'étude du rôle éventuel de connaissances sur les régularités phonologiques de la langue dans les processus de segmentation de la parole en mots. La nécessité d'un contrôle plus précis du matériel linguistique utilisé a été mise en évidence dans le cadre d'une analyse de la fréquence des groupes de consonnes dans un lexique informatisé de la langue française. Cette étude, présentée dans le Chapitre 4, permet de mettre en évidence un lien étroit entre la structure syllabique des groupes de deux consonnes et leur fréquence d'occurrence dans les mots de la langue. Les groupes consonantiques tautosyllabiques s'avèrent en général plus fréquents que les groupes hétérosyllabiques. Cette observation conduit à envisager diverses interprétations possibles des effets mis en évidence par McQueen (1998) et par Vroomen & De Gelder (1999). Trois axes interprétatifs ont été proposés à partir de cette observation. Le premier se rapporte au modèle COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987). Un déséquilibre de la fréquence des séquences de phonèmes comparées risque d'engendrer également un déséquilibre du nombre de candidats lexicaux qui sont maintenus dans la cohorte lorsque le phonème contextuel a été identifié. Cette interprétation a été mise à l'épreuve des données expérimentales dans les expériences 1 et 2 (Chapitre 5). Les deux autres interprétations proposées sont directement liées à la fréquence des séquences de phonèmes. Deux classes de modèles peuvent en effet rendre compte des effets observés si, dans

le matériel utilisé, structure syllabique / phonotactique et fréquence d'occurrence sont liées. Dans le cadre d'un modèle probabiliste de la segmentation lexicale, les suites fréquentes ont tendance à être regroupées à l'intérieur d'une hypothèse lexicale alors que les suites rares favorisent l'insertion d'une frontière de mots. Si ce modèle nécessite le recours à une conception prélexicale des processus de segmentation, il n'est nullement nécessaire de supposer que les informations utilisées par le système cognitif pour savoir si une suite est rare ou fréquente sont d'ordre phonologique. Le recouvrement des distributions de probabilité d'occurrence décrit dans la section 2.2.2 du Chapitre 4 confirme cette indépendance des représentations phonologiques et probabilistes. Par ailleurs, la fréquence des suites de phonèmes détermine également le nombre de candidats lexicaux qui seront activés dans le cadre d'un modèle à alignement exhaustif (McClelland & Elman, 1986 ; Norris, 1994). Cette classe de modèles peut donc prédire les résultats observés, la sélection du candidat approprié se faisant parmi une quantité de compétiteurs plus importante lorsqu'une séquence de phonèmes est fréquente que lorsqu'elle est rare. L'effet observé serait alors, en réalité, localisé à un niveau lexical de traitement. Le recouvrement des distributions observé dans le Chapitre 4 fournit la possibilité de dissocier structure phonologique et fréquence pour la comparaison de diverses catégories de groupes consonantiques. Ce second groupe d'interprétations a été testé dans les expériences 3 à 6 (Chapitre 6).

1. Résumé des résultats obtenus

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

1.1. Effets observés

1.1.1. *Contraintes phonologiques et sélection lexicale*

Les expériences du Chapitre 4 (notamment l'Expérience 1) permettent d'affirmer que, même en l'absence d'un déséquilibre dans la taille de la cohorte de candidats, il est possible de mettre en évidence un effet de l'alignement entre frontière phonologique / probabiliste et frontière lexicale. Les latences de détection observées dans la tâche de word-spotting sont plus longues lorsque l'occlusive finale du mot est suivie d'une liquide que lorsque elle est suivie

d'une fricative ou d'une occlusive. Cet effet se restreint cependant à une partie du matériel ; en effet l'analyse par item ne s'avère significative que si l'on se restreint aux mots de l'échantillon qui ont le nombre de candidats le plus élevé. Un critère lexical détermine donc l'intervention de processus censés intervenir au cours des étapes prélexicales du traitement. Cette incongruité conduit à accentuer l'idée d'une interprétation lexicale de l'effet.

1.1.2. *Régularités phonologiques, fréquence et compétitions lexicales*

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

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

1.2. Discussion générale

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

expériences sont présentées. La suite du chapitre constitue une réflexion sur les résultats obtenus dans ce travail de thèse.

1.2.1. *Processus de traitement précoces ou stratégies ?*

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

Nous avons déjà abordé la problématique du rôle des stratégies de réponse lors des discussions concernant les expériences 3 et 5 et avons envisagé que le développement d'attentes ait pu masquer l'émergence de l'effet attendu. Les stratégies de réponse développées par les participants constituent typiquement une classe de processus décisionnels, lesquels ne sont mis en œuvre qu'une fois l'intégralité des traitements effectuée. L'expérience 6, dans laquelle un effet significatif du contexte phonologique est mis en évidence pour les mots les plus fréquents de l'échantillon, a été mise en place afin de limiter l'influence de stratégies de réponse sur les mesures effectuées. La mise en évidence de l'effet dans cette expérience devrait donc contribuer à invalider l'idée d'un effet stratégique. Il est cependant possible que l'utilisation même de procédures de segmentation syllabiques / phonotactiques soit déterminée par un niveau optimal de difficulté de la tâche. Ainsi, lorsque la tâche est trop difficile - ceci semble notamment être le cas lorsque la contribution de la fréquence est neutralisée et que l'ordre de présentation des stimuli de chaque condition est aléatoire - les réponses des participants seraient perturbées ; ce qui conduirait à masquer l'effet phonologique. Lorsque la tâche des participants est facilitée par

un regroupement des conditions de contexte en début et en fin de liste, cette succession séquentielle des conditions expérimentales limiterait l'influence des réponses précédentes sur la réponse en cours et favoriserait l'émergence de l'effet. Le paradigme de word-spotting reste néanmoins particulièrement difficile quelle que soit la procédure employée -preuve en sont les taux d'erreurs élevés dans toutes les expériences qui sont présentées ici. Il est légitime de se demander si cet effet se maintiendrait dans une tâche plus facile. Si l'utilisation de procédures de segmentation phonologique n'est mise en place que pour compenser une difficulté particulière à isoler les mots dans le matériel, l'effet disparaîtrait probablement. Si au contraire, ces procédures sont réellement mises en œuvre par le système cognitif lorsqu'il traite un signal de parole -que ce soit dans une situation naturelle ou expérimentale- l'effet devrait persister avec une tâche plus facile.

1.2.2. *Lien entre régularités phonologiques et propriétés lexicales*

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

L'effet du nombre de compétiteurs d'un mot peut être implémenté à deux niveaux de représentation. Vitevitch, Luce, Pisoni, & Auer (1999 ; cf. aussi Luce, 1998 ; Vitevitch & Luce,

1999 pour une présentation du modèle NAM⁴³ sur lequel est fondé ce travail) présentent un modèle de l'implémentation des paramètres liés au nombre de voisins phonologiques d'un mot. Au niveau lexical, un mot ayant beaucoup de voisins phonologiques, c'est à dire pour lequel il existe beaucoup d'autres mots dans le lexique qui partagent 1 à N phonèmes avec ce mot, entrera en compétition avec un nombre élevé de candidats lexicaux. Un accroissement du nombre de voisins lexicaux rendra sa reconnaissance plus difficile puisqu'il faut alors sélectionner le candidat correct parmi une quantité plus élevée de mots possibles. Au niveau prélexical, on peut prédire un effet inverse. Si un mot a beaucoup de voisins lexicaux, les diphtonges dont il est constitué sont nombreux (puisque présents dans une quantité importante de mots). Cette fréquence pourra alors faciliter l'identification des phonèmes qui le constituent (cf. le rôle des probabilités transitionnelles et leur rapport avec le paradigme de la *théorie de l'information*, section 2.1.3 du Chapitre 4) et, de fait, l'identification des mots. Observer une interaction entre des variables respectivement prélexicale et lexicale peut donc s'expliquer en réalité par une interaction faisant intervenir deux variables prélexicales... ou deux variables lexicales. En ce qui concerne la restriction de l'effet du contexte phonologique à un ensemble de mots partageant des caractéristiques lexicales (le nombre de compétiteurs), on peut donc imaginer que ce que nous avons considéré jusqu'à maintenant comme une variable lexicale désigne peut-être aussi une spécificité des mots de notre échantillon qui pourrait en fait s'exprimer en termes sublexicaux. Il est donc intéressant d'analyser plus en détails les caractéristiques du matériel utilisé pour interpréter la restriction de l'effet à un sous-ensemble des mots de l'échantillon.

1.2.3. *Mode d'intégration des processus de segmentation*

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

⁴³ *Neighbourhood Activation Model.*

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

2. Niveau d'implémentation

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

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

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

2.1.1. Conditions d'émergence de l'effet phonologique

Lorsque l'on compare les données obtenues pour des groupes tautosyllabiques et hétérosyllabiques de même fréquence, une alternance aléatoire des stimuli de chacune des conditions ne semble pas permettre de faire émerger cet effet. Malgré l'accroissement du nombre de stimuli de remplissage effectué dans l'expérience 5, lequel était censé limiter l'influence des réponses précédentes sur la réponse en cours, aucun effet même minime n'est observé. Or si l'on regroupe les stimuli de chaque condition en début et en fin de liste, l'effet d'alignement attendu se manifeste. Ce changement des résultats lié à un changement de procédure peut s'expliquer par une stabilisation de la structure syllabique ou phonotactique des stimuli-test ; lorsque les stimuli

sont mélangés aléatoirement, l'occurrence d'un stimulus CV#CCVC peut induire les participants à attendre un stimulus dont la structure est identique. Si le stimulus suivant correspond au contraire à une structure CVC#CVC, la discordance entre les attentes des participants et la structure effective du stimulus peut conduire à un accroissement de la latence de détection ou à une réponse erronée alors que cette structure est supposée donner lieu à des latences plus courtes et à une quantité moins élevée d'erreurs. Dans la situation inverse, une séquence CVC#CVC suivie d'un stimulus CV#CCVC devrait contribuer à accentuer l'effet attendu. Cependant, la combinaison de ces deux formes d'alternance peut masquer l'émergence de l'effet. Lorsque les conditions sont regroupées en début et en fin de liste, ces phénomènes d'attente sont neutralisés par la régularité de structure des stimuli-test présentés. Il est important de noter que cette procédure aurait tout aussi bien pu conduire à une suppression de l'effet puisque cette situation aurait pu permettre aux participants de prédire la structure des stimuli dans chaque demi-liste. Au contraire, le choix de cette procédure contribue à l'émergence de l'effet.⁴⁴

2.1.2. *Caractère écologique des processus impliqués*

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

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

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

Si la difficulté de la tâche incite les participants à utiliser ce découpage syllabique / phonotactique, rien ne permet d'affirmer que cette procédure est effectivement mise en œuvre dans les situations naturelles de traitement de la parole. Si le découpage phonologique effectué par les participants dans l'expérience 6 n'est que le fruit d'une stratégie de traitement destinée à réussir une tâche difficile, le recours à des connaissances sur la structure phonotactique ou syllabique des séquences de parole ne peut être considéré que comme un processus stratégique post-perceptif qui n'intervient qu'après les processus de traitement de l'information. Si au contraire ce processus est caractéristique des traitements naturels, c'est à dire s'il est effectivement mis en place dans le cadre du traitement de l'information acoustique, phonétique et lexicale, il devrait pouvoir être mis en évidence dans une tâche plus facile que le *word-spotting*. Nous avons donc décidé de reproduire les expériences 3 et 4 en changeant de paradigme expérimental. Vroomen & De Gelder (1999) avaient observé cet effet d'alignement à l'aide d'une tâche de détection de phonèmes. Or cette tâche donne lieu à des latences de détection beaucoup plus courtes et à des taux d'erreur plus faibles que la tâche de *word-spotting*. C'est donc tout naturellement que nous avons choisi la détection de phonème pour étudier à

nouveau la contribution respective de la fréquence et de la structure phonologique dans les processus de segmentation lexicale. Deux expériences ont été conduites avec les mêmes participants. Ceux-ci n'avaient pas pris part aux précédentes. La première est une réplique de l'expérience 3 ; des groupes tautosyllabiques et hétérosyllabiques de même fréquence sont comparés. La seconde reprend les objectifs de l'expérience 4 avec une comparaison de groupes fréquents et rares étant tous hétérosyllabiques.

2.1.3. *Expérience 7 : Groupes de consonnes de même fréquence*

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

2.1.3.1. Méthode

2.1.3.1.1. *Sujets*

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

2.1.3.1.2. *Matériel*

Le matériel linguistique change notablement par rapport aux stimuli utilisés dans la tâche de word-spotting. Ce sont des groupes nominaux constitués d'un déterminant (l'adjectif indéfini 'un' ou 'une'), d'un nom commun monosyllabique et d'un adjectif qualificatif constitué de deux ou trois syllabes (par exemple 'un sac lavable'). Nous inspirant du travail de Vroomen & De Gelder (1999), nous avons choisi de conduire cette expérience avec des suites de deux mots. Il eut certainement été possible de réutiliser le matériel de word-spotting en changeant la tâche des participants. Il était cependant intéressant d'utiliser un matériel se rapprochant des stimuli couramment rencontrés dans les énoncés de parole naturels. Puisque la problématique que nous souhaitons aborder ici est celle de la distinction entre processus de traitement et stratégies post-

lexicales, ceci permet d'accroître la proximité avec une situation naturelle. Ce choix apporte cependant des contraintes nouvelles pour la constitution du matériel. En effet, il convient de sélectionner des suites nom commun + adjectif qui ne soient ni incongrues ni trop prédictibles. Huit mots ont chacun été appariés avec deux adjectifs post-posés. La consonne finale du nom commun est toujours l'occlusive /k/. Associée à l'occlusive finale du nom commun, la consonne initiale de l'adjectif qualificatif constitue un groupe tautosyllabique (lorsque la consonne initiale de l'adjectif est une liquide, par exemple /kl/ dans 'un sac lavable', /œsaklavablə/) ou hétérosyllabique (pour les consonnes fricatives ou occlusives, par exemple /kt/ dans 'un sac tissé', /œsaktise/). Ces groupes sont appariés sur la base de leur probabilité d'occurrence brute. Aucun des stimuli-test ne contient une voyelle qui permettrait d'influencer le rattachement syllabique de la consonne suivante (comme /o/ ou /e/). Quatre groupes de consonnes différents ont été comparés (cf. Tableau 17).

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

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

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

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

2.1.3.1.3. Procédure

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

2.1.3.2. Résultats

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

2.1.3.2.1. Temps de réaction

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

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

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

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

2.1.3.2.2. Taux d'erreur

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

2.1.3.3. Discussion

Les données de l'expérience 7 reproduisent l'absence d'effet déjà observée dans les expériences 3 et 5. Contrairement à ce qui avait été observé dans l'expérience 6, le recours à la tâche de détection de phonèmes plutôt qu'à celle de *word-spotting* ne permet pas de mettre en évidence un effet de l'alignement entre frontières phonologique et lexicale. Ce changement de tâche a été effectué afin de répondre à la question essentielle du niveau d'intervention des connaissances sur les contraintes phonologiques de la langue dans les processus de segmentation de la parole en mots. Nous avons formulé l'hypothèse que si ces représentations phonologiques sont utilisées au cours des étapes précoces du traitement, l'effet devrait émerger dans n'importe quelle situation expérimentale *ad hoc*. Au contraire, si l'effet observé dans l'expérience 6 peut relever de processus décisionnels, l'adoption d'une tâche beaucoup plus facile aurait du donner lieu à une disparition de l'effet. La comparaison des taux d'erreurs mesurés dans les précédentes expériences de *word-spotting* et dans la tâche de détection de phonèmes présentée confirme la différence de difficulté des deux tâches. En *word-spotting*, on observe des taux d'erreurs qui

oscillent entre 15.0% (Expérience 1) et 23.9 % (Expérience 5). La proportion d'erreurs observée dans l'expérience 6 est nettement plus faible (5.2 %). L'adoption d'une tâche moins difficile que le *word-spotting* semble donc contribuer à l'abandon de procédures de segmentation phonologique de la chaîne de parole ; ce qui semble confirmer leur caractère artificiel. Il convient cependant, avant de conclure à une implémentation post-perceptive de l'effet phonologique observé dans l'expérience 6, d'acquérir la certitude que des effets mis en évidence avec le paradigme de *word-spotting* peuvent être répliqués dans une tâche de détection de phonèmes. Dans l'expérience 4, la comparaison de groupes de consonnes hétérosyllabiques fréquents et rares donne lieu à des taux d'erreurs qui diffèrent significativement dans l'analyse par sujet. Cet effet se manifeste malgré une procédure tout à fait similaire à celle de l'expérience 3 dans laquelle aucun effet n'émerge ; il est donc judicieux de répliquer l'expérience 4 avec un matériel similaire.

2.1.4. *Expérience 8 : Estimation du rôle de la fréquence*

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

2.1.4.1. Méthode

2.1.4.1.1. Sujets

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

2.1.4.1.2. Matériel

Les caractéristiques du matériel linguistique sont similaires à celles des stimuli de la précédente expérience. Ce sont des groupes nominaux constitués d'un déterminant (l'adjectif indéfini 'un' ou 'une'), d'un nom commun monosyllabique et d'un adjectif qualificatif constitué de deux ou trois syllabes. Douze mots ont chacun été appariés avec deux adjectifs post-posés. La consonne finale du nom commun est toujours l'occlusive /p/ ; c'est la consonne-cible. Associée à l'occlusive finale du nom commun, la consonne initiale de l'adjectif qualificatif génère un groupe fréquent (par exemple /ps/ dans 'une coupe soldée', /ynkupsɔlde/) ou rare (par exemple /pʃ/ dans 'une coupe chinoise', /ynkupʃinwaz/). Ces groupes sont tous hétérosyllabiques. Aucun

des stimuli-test ne contient une voyelle qui permettrait d'influencer le rattachement syllabique de la consonne suivante (comme /o/ ou /e/). Quatre groupes de consonnes différents ont été comparés (cf. Tableau 19).

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

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

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

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

2.1.4.1.3. Procédure

Les participants avaient pour consigne d'appuyer le plus rapidement possible sur un bouton réponse dès qu'ils identifiaient le phonème-cible. Un seul phonème devait être détecté dans cette expérience (/p/). Il pouvait apparaître dans n'importe quelle position mais seules les réponses

fournies pour le phonème final du mot monosyllabique sont étudiées. Les temps de réaction étaient mesurés depuis le début du stimulus et corrigés automatiquement par le logiciel de passation afin de refléter la latence entre l'occlusion de la consonne cible et la réponse du sujet. Avant la phase expérimentale proprement dite, les participants se familiarisaient avec la tâche à l'aide d'un matériel différent.

2.1.4.2. Résultats

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

2.1.4.2.1. Temps de réaction

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

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

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

Cette différence est significative dans l'analyse par sujet ($F_1(1,23) = 6.371$, $p < .05$) ainsi que dans celle par item ($F_2(1,11) = 10.716$, $p < .01$). La détection du phonème final d'un mot monosyllabique est donc plus lente lorsque celui-ci est prononcé au début d'un groupe fréquent que lorsqu'il forme, avec le phonème contextuel, un groupe relativement rare dans la langue. Cet

effet est tout à fait similaire à celui qui avait été observé dans une tâche de word-spotting -uniquement dans l'analyse par sujet- avec un matériel semblable (expérience 4).

2.1.4.2.2. Taux d'erreur

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

2.1.5. Discussion générale

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

de processus de segmentation fondés sur des représentations phonologiques plus grandes que le phonème.

Le cœur du travail présenté ici consiste à poser la question du rôle éventuel d'informations dont le support acoustique excède le phonème dans les processus de perception de la parole. Les régularités phonologiques étudiées sont liées à l'agencement des phonèmes entre eux. Que l'on parle de régularités phonotactiques ou de structure syllabique, ce sont les caractéristiques des *séquences* de phonèmes qui déterminent leur statut. Une suite de phonèmes constitue une *bonne forme* en début de mot, est prononcée en *attaque* syllabique, en fonction d'une certaine relation entre les segments phonémiques dont elle est composée. Dans les études présentées, les caractéristiques d'un segment ne déterminent pas son rattachement à l'arborescence syllabique ou son degré d'acceptabilité en début de mot⁴⁵. Les seules représentations valides pour l'implémentation de procédures de segmentation de la parole fondées sur ce type d'informations reposent sur les régularités de *co-occurrence* des segments qui les constituent. Or la tâche de détection de phonèmes peut tout à fait être réalisée sans faire intervenir ce type de connaissances contextuelles. La tâche des participants consiste essentiellement, à partir d'une cible prédéterminée, à identifier un phonème -donc un *segment*- dans le signal acoustique présenté. Dans la tâche de *word-spotting* au contraire, l'information à extraire est constituée de plusieurs phonèmes ; l'objectif primordial étant de dissocier deux ensembles de phonèmes de manière à ce que l'un d'entre eux constitue un mot de la langue. Alors que la tâche de détection de phonèmes implique l'accès à des représentations phonémiques, celle de *word-spotting* ne nécessite pas d'identifier -en tout cas consciemment- les phonèmes qui constituent ce signal de parole. Plus précisément, il n'est pas certain que le type de représentation phonémique développé dans la tâche de détection de phonèmes soit identique à celui qui se met en place lorsque l'on traite un signal de parole afin d'identifier les mots dont il est constitué (cf. Foss & Blank, 1980 ; Segui et al., 1981 ; Szende, 1980 pour des discussions plus approfondies sur ce thème). L'on est alors en mesure d'envisager que les informations sur lesquelles repose la décision de détection du phonème-cible ne sont pas nécessairement influencées par des représentations suprasegmentales. La tâche de détection de phonème pourrait alors être considérée comme *moins écologique* que celle de *word-spotting* ; ce qui expliquerait l'absence d'effet d'alignement entre frontières suprasegmentale et lexicale. Les données présentées par Vroomen & De Gelder (1999) avec une

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



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

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

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

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

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

2.2. Interaction entre variables prélexicales et lexicales

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

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

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

2.2.1. *Nombre de voisins lexicaux et fréquence des diphtonges*

Vitevitch, Luce, Pisoni, & Auer (1999) montrent que les mots qui ont une densité de voisinage lexical élevée⁴⁶ sont souvent constitués de suites de phonèmes fréquentes dans la langue. La variable que nous avons utilisée pour dichotomiser le matériel de l'expérience 1 n'est pas le 'nombre de voisins lexicaux' tel qu'il est décrit par Vitevitch, Luce, Pisoni, & Auer (1999). Lorsque nous parlons ici de 'nombre de compétiteur', c'est à partir d'un calcul prenant en compte le nombre de mots *commençant* par la séquence de phonèmes d'un mot donné. Le 'nombre de voisins lexicaux' correspond au contraire à la quantité de mots du lexique qui partagent un ou plusieurs phonèmes, et ce quelle que soit leur position dans les mots. Ces deux méthodes de calcul sont tout aussi pertinentes l'une que l'autre. Dans le cas d'un modèle comme COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987), c'est le 'nombre de compétiteurs' tel que nous l'avons calculé qui influence les processus de reconnaissance. Des modèles comme TRACE (McClelland & Elman, 1986) ou le *Neighborhood Activation Model* (Luce & Pisoni, 1998) font au contraire intervenir des processus qui reposent essentiellement sur le 'nombre de voisins lexicaux' puisque les candidats activés peuvent l'être en tout endroit de la

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

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

Si l'influence du nombre de compétiteurs est envisagée sous son angle lexical, il conviendrait de considérer l'effet du contexte phonologique observé dans l'Expérience 1 comme la manifestation d'une effet dont la source serait à rechercher au niveau du lexique (cf. le lien entre fréquence et compétitions lexicales) ou à une étape post-perceptive / décisionnelle. Si au contraire l'influence du nombre de compétiteurs sur l'émergence de l'effet de contexte s'explique par des caractéristiques sublexicales des mots, il serait possible d'envisager cette restriction de l'effet observé dans la première expérience comme une interaction entre deux variables intervenant au cours des étapes prélexicales du traitement. Cette seconde solution constituerait donc un argument en faveur d'une influence purement prélexicale des effets phonologiques observés. Nous avons donc cherché à identifier des régularités structurelles des items de l'Expérience 1 qui correspondraient à la dissociation que nous avons opérée entre les mots qui ont peu ou beaucoup de compétiteurs. Cette régularité existe et se manifeste dans la structure consonne-voyelle des mots utilisés. Dans le matériel que nous avons choisi, tous les mots se terminaient par une séquence VC. Nous avons cependant mélangé des mots *commençant* par une consonne unique (CVC) avec d'autres dont l'initiale formait un groupe de consonnes (CCVC). Les mots qui ont beaucoup de compétiteurs ont presque tous une structure CVC (16 sur 18) alors que c'est l'inverse pour ceux qui ont peu de compétiteurs (4 CVC et 9 CCVC). Rappelons que les suites CC sont nettement plus rares dans la langue que les consonnes intervocaliques (cf. Chapitre 4). Il est donc possible que nous ayons proposé une interprétation

lexicale de la restriction de l'effet alors qu'elle est explicable en termes probabilistes donc prélexicaux ; deux variables prélexicales se combineraient alors, l'une reposant sur la fréquence des diphones à l'intérieur des mots et l'autre sur le statut phonologique du groupe de consonnes médian. Dans l'expérience 6, c'est la restriction fondée sur la fréquence d'usage des mots qui contribue à faire émerger un effet stable du type de groupe consonantique. Une analyse du matériel des expériences 3, 5 et 6⁴⁷ n'a cependant pas permis de mettre au jour une régularité similaire entre les paramètres lexicaux liés à la fréquence des mots de l'échantillon et des régularités sublexicales comme la fréquence des diphones. La structure Consonne-Voyelle des mots de l'échantillon est équitablement distribuée entre les mots les plus fréquents (6 CVC, 3 CCVC) et les plus rares (5 CVC, 3 CCVC) de l'échantillon. Les données de cette expérience ne permettent pas de confirmer les conclusions avancées par l'étude de la structure phonémique du matériel de l'expérience 1. Cette expérience nous laisse donc dans l'incertitude concernant le lien entre la variable lexicale que constitue la fréquence des mots et la variable sublexicale supposée à l'œuvre dans la segmentation phonologique du signal de parole. Il est possible que nous n'ayons pas su trouver un paramètre sublexical adéquat dans le matériel de l'expérience 6. Il est aussi possible que l'on soit effectivement en présence d'une restriction de l'effet fondée sur des informations lexicales ; ce qui inciterait à concevoir l'effet d'alignement observé dans l'expérience 6 comme un effet post-perceptif influencé par les traitements lexicaux. Cette conclusion serait conforme à l'absence d'effet d'alignement entre frontières phonologique et lexicale dans la tâche de détection de phonèmes. Une autre interprétation se doit cependant d'être mentionnée. Pour cela, il convient d'aborder une réflexion sur l'activité cognitive des auditeurs lorsqu'ils traitent un signal de parole et, plus particulièrement, lorsqu'ils réalisent une tâche expérimentale.

2.2.2. *Intégration des informations*

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

⁴⁷ Pour lesquelles les stimuli-test sont les mêmes.

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

On peut au contraire envisager que ces deux processus sont relativement indépendants. Le décours temporel des activations lexicales pourrait s'effectuer sans qu'aucune modification ne soit introduite par le processus de segmentation phonotactique ou probabiliste. Par contre, ces informations pourraient être utilisées comme une source d'information supplémentaire pour décider de la présence d'un mot dans le stimulus. On sait que certains mots sont plus faciles à reconnaître que d'autres. Diverses caractéristiques aussi bien lexicales que prélexicales peuvent influencer la durée nécessaire ou le rapport Signal/Bruit acceptable pour être en mesure d'identifier un mot. Ces caractéristiques peuvent être mises en évidence dans des tâches comme la décision lexicale ou l'identification de mot dans du bruit, paradigmes dans lesquels le découpage lexical est déjà présent puisque ce sont des mots isolés qui sont présentés. En fonction de la difficulté intrinsèque des mots pour être reconnus comme présents dans le lexique, cette difficulté reposant sur la fréquence d'usage, le nombre de voisins lexicaux, le nombre de candidats maintenus dans la cohorte ou la fréquence des diphtonges qui les constituent, l'utilisation de procédures de segmentation phonotactique ou syllabique pourrait n'être pas efficace. On peut par exemple supposer que si des mots sont très faciles à reconnaître, la mise en œuvre de connaissances sur les régularités de la langue sera -dans une tâche de *word-spotting*- de peu d'intérêt pour localiser la frontière entre le mot et le reste du non-mot. Les informations utilisées pour identifier le mot en isolation pourraient suffire. Au contraire, des mots intrinsèquement difficiles à identifier pourraient bénéficier d'indices supplémentaires pour que la frontière entre

le mot et le reste du non-mot soit localisée. Cette explication pourrait rendre compte de la restriction de l'effet aux mots qui ont beaucoup de compétiteurs dans l'expérience 1. Le modèle NAM (Luce & Pisoni, 1998) prédit que les mots qui ont beaucoup de voisins lexicaux seront plus difficiles à identifier parce qu'ils entrent en compétition avec une quantité plus élevée de candidats. Selon notre proposition, cette difficulté permettrait au système de recourir à des processus de segmentation phonotactique ou syllabique alors que ces mêmes processus n'influenceraient pas les latences de décision ou les taux d'erreur lorsque le mot entre en compétition avec une faible quantité de candidats. On retrouve ici des propositions similaires à celles qui étaient faites dans le cadre du modèle RACE (Cutler & Norris, 1979). Dans ce modèle (cf. 2.1 du Chapitre 1), les informations lexicales n'influencent les processus d'identification phonémique que si l'accès au lexique est 'plus rapide' que l'accès à une représentation phonémique abstraite qui prendrait appui sur une voie non-lexicale⁴⁸. La restriction de l'effet aux mots fréquents (expérience 6) semble plus difficile à expliquer. Les mots fréquents sont en effet plus faciles à identifier que les mots rares. Ceci n'est cependant vrai que toutes choses égales par ailleurs. Or l'on peut tout à fait envisager que la distinction entre mots fréquents et rares dans notre échantillon correspond à une autre distinction⁴⁹, par exemple le nombre de 'voisins lexicaux'. Si les mots fréquents ont plus de voisins lexicaux que les mots rares, l'effet du nombre de voisins lexicaux devrait contribuer, au contraire de la fréquence, à rendre plus difficile leur identification. Tout dépend alors de l'interaction entre ces deux variables. Si la fréquence ne suffit pas à compenser l'effet du nombre de voisins lexicaux, les mots fréquents risquent d'être plus difficiles à identifier que les mots rares ; ce qui permettrait de fournir la même explication à la restriction de l'effet dans l'expérience 6.

3. Mode d'intégration

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

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

conceptions du lien entre les étapes prélexicale et lexicale du traitement de la parole. Le premier cadre théorique découle directement du paradigme⁵⁰ connexionniste et fait reposer l'émergence de l'effet d'alignement sur le niveau d'activation des unités lexicales. Le second repose sur des propositions et des observations diverses dans le champ de la psychologie cognitive, que ce soit dans le cadre spécifique de la psycholinguistique ou dans les domaines plus éloignés de l'étude des processus perceptifs tels que la vision ou le traitement musical.

3.1. Approches théoriques

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

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

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

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

⁵⁰ Au sens de *paradigme théorique*.

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

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

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

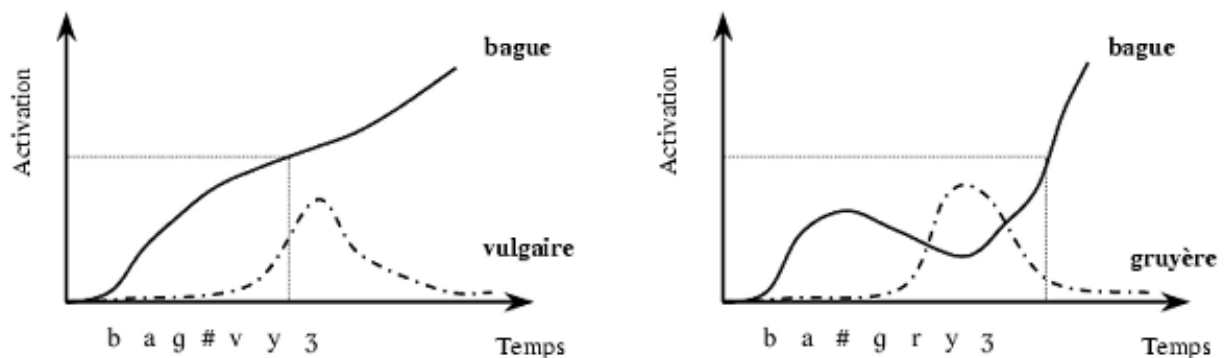


Figure 31 : Prédiction (imaginaire) du comportement d'un modèle dans lequel une procédure similaire à la Metrical Segmentation Strategy (Norris et al., 1995) accentuerait le niveau d'activation des unités lexicales qui sont alignées avec les débuts de syllabe. Le mot à reconnaître est 'bague' en position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

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

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

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

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

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

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

3.1.2.1. Intégration des informations perceptives

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

3.1.2.1.1. *Le rôle des alternances prosodiques en français*

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

concevoir ces effets liés à la structure prosodique des énoncés comme dépendant de procédures de traitement qui ne seraient pas spécifiques à la langue, pas plus qu'au langage, mais reposeraient sur des procédures générales de traitement.

3.1.2.1.2. *L'organisation hiérarchique des pièces musicales*

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

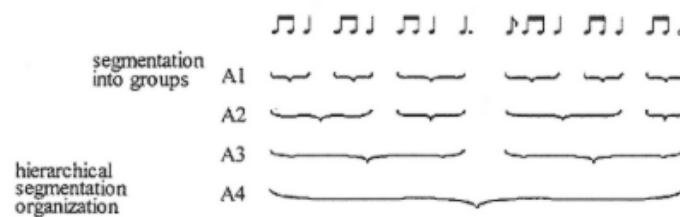


Figure 32 : Divers niveaux d'organisation hiérarchique dans la perception musicale (extrait de Drake, 1998)

Ces processus d'organisation hiérarchique des pièces musicales peuvent être mis en rapport avec les processus de segmentation du signal de parole fondés aussi bien sur les alternances prosodiques (Banel & Bacri, 1994 ; Cutler & Norris, 1988) que sur les caractéristiques des séquences de phonèmes (McQueen, 1998 ; Vroomen & de Gelder, 1999). En français, l'allongement de la durée d'une syllabe (mais aussi l'élévation de l'intensité ou de la F_0) signale la fin d'un groupe prosodique. Ces groupes prosodiques peuvent être assimilés à des groupements perceptifs correspondant à un niveau hiérarchique relativement large (2 syllabes ou plus). De même, l'occurrence d'un groupe consonantique phonotactiquement illégal permettrait de signaler une frontière de groupement à partir d'indices non plus rythmiques mais phonotactiques. Ces groupements pourraient correspondre à une seule syllabe. Parallèlement, l'occurrence d'une suite VCV induirait une segmentation syllabique fondée sur le statut d'attaque de la consonne et correspondrait également à des groupes hiérarchiques monosyllabiques. Il est essentiel de noter que des processus tout à fait similaires peuvent être

décrits dans ces deux domaines d'étude. Il est même possible d'envisager que ces processus trouvent leur source dans des procédures de traitement cognitif communes et que les effets mis en évidence dans les études sur la perception de la parole peuvent reposer sur des routines de segmentation qui ne sont pas spécifiques du traitement du langage parlé (c'est ce qui est proposé par Banel & Bacri, 1994). Le modèle proposé par Cutler & Norris (1988) ne peut en tout cas pas rendre compte des effets observés en français. Puisque les groupes prosodiques français se terminent par une syllabe accentuée, il n'est pas possible de faire reposer le déclenchement des activations lexicales (ou leur modification) sur l'alignement entre syllabe accentuée et début de mot⁵².

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

3.1.2.1.3. Perspectives dérivées de la théorie de la Gestalt

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

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



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

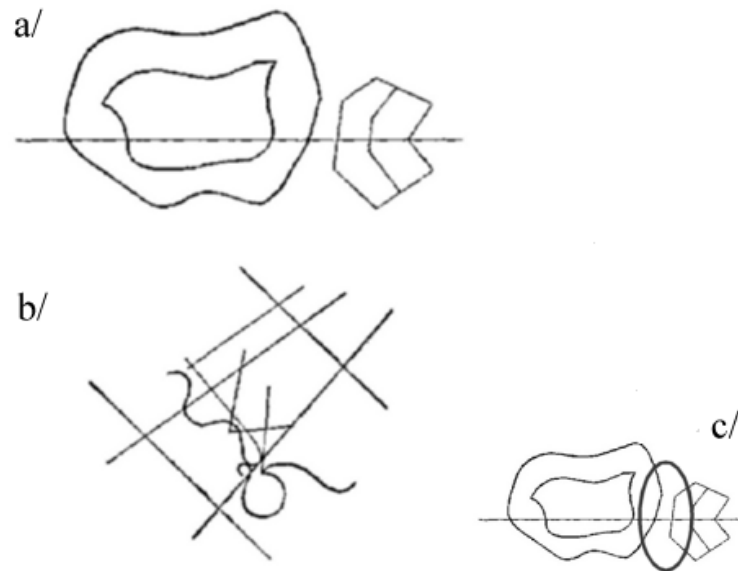


Figure 33 : Exemple de segmentation perceptive liée au phénomène de continuité des contours (ou *loi de bonne continuité*). En b/, le chiffre '4' est nettement dissocié des autres lignes. En a/, il est particulièrement difficile de percevoir ce chiffre à cause de la continuité entre les lignes du chiffre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

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

3.1.2.2. Segmentation hiérarchique et niveaux d'activation lexicale

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

effet.

notamment) mais aucun de ces sous-processus de traitement n'influencerait l'autre. Ces informations seraient comparées à un moment du traitement que nous ne pouvons déterminer ici -par exemple au niveau de la prise de décision- et la réponse du système dépendrait de la comparaison des deux catégories d'informations⁵³.

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

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

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

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

3.1.3. *Une expérience qui permettrait de tester cette hypothèse*

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

traitement du langage parlé, il devrait donc être possible de mettre en évidence le même type d'effets.

Résumé

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