Rôle de la coarticulation dans la reconnaissance des mots

Noël Nguyen

L'année psychologique

Rôle de la coarticulation dans la reconnaissance des mots

Noël Nguyen

Abstract

Summary : The role of coarticulation in word recognition.

Recent experiments dealing with the role of coarticulatory effects in word recognition are reviewed. In contrast to a longstanding view, these studies tend to show that lexical access is not based upon an abstract phonological representation ofthe speech signal. Perceptual data collected using different experimental paradigms (gating, lexical decision, cross-modal priming, etc.) show on the contrary that listeners may be directly sensitive to small coarticulatory eues in the identification ofwords. The input to the lexical search seems therefore to be a detailed phonetic representation incorporating information about the fine-grained acoustic structure of speech. Implications for current models of word recognition are discussed.

Key words : coarticulation, word recognition, speech perception.

Résumé

Résumé

Dans cet article, nous passons en revue un ensemble d'expériences récemment réalisées sur le rôle des effets de coarticulation dans la reconnaissance des mots. En désaccord avec une hypothèse qui a longtemps prévalu, ces travaux donnent à penser que la reconnaissance d'un mot ne s'opère pas à partir d'une représentation phonologique abstraite du signal de parole. Les données les plus récentes font au contraire apparaître que l'auditeur peut, dans l'identification d'un mot, être directement influencé par une grande variété de contrastes phonétiques extrêmement ténus. Les implications de ces résultats en ce qui concerne les modèles actuels de la reconnaissance des mots sont discutées.

Mots-clés : coarticulation, reconnaissance des mots, perception de la parole.

L'Année psychologique, 2001, *101*, 125-154

Laboratoire Parole et Langage, CNRS
Département de phonétique
Université de Provence, Aix-Marseille I[1]

RÔLE DE LA COARTICULATION DANS LA RECONNAISSANCE DES MOTS

par Noël NGUYEN[2 3]

SUMMARY : The role of coarticulation in word recognition.

Recent experiments dealing with the role of coarticulatory effects in word recognition are reviewed. In contrast to a longstanding view, these studies tend to show that lexical access is not based upon an abstract phonological representation of the speech signal. Perceptual data collected using different experimental paradigms (gating, lexical decision, cross-modal priming, etc.) show on the contrary that listeners may be directly sensitive to small coarticulatory cues in the identification of words. The input to the lexical search seems therefore to be a detailed phonetic representation incorporating information about the fine-grained acoustic structure of speech. Implications for current models of word recognition are discussed.

Key words : coarticulation, word recognition, speech perception.

INTRODUCTION

Le signal de parole ne forme pas une séquence linéaire de segments indépendants les uns des autres. De multiples travaux ont montré en particulier l'importante variabilité présentée par les sons de la parole selon leur entourage phonétique (Perkell et Klatt, 1986). Cette variabilité a été en partie attribuée au fait que les mouvements accomplis par les articulateurs dans la production de la parole se chevauchent sur l'axe temporel (Öhman, 1966 ; Fowler, 1980 ; Browman et Goldstein, 1986). Dans une syllabe de

1. 29, avenue Robert-Schuman, 13621 Aix-en-Provence Cedex 1.
2. E-mail : noel.nguyen@lpl.univ-aix.fr.
3. Je remercie Cécile Fougeron et Uli Frauenfelder pour l'aide qu'ils m'ont apportée.

type CV, par exemple, pour citer le cas sans doute le plus étudié, les gestes articulatoires associés à la consonne initiale et à la voyelle qui la suit sont partiellement superposés. Ces phénomènes de recouvrement temporel, généralement désignés sous le terme de coarticulation, font de la relation entre unités phonétiques et signal acoustique de sortie une relation non biunivoque : chaque portion du signal est le plus souvent à mettre en relation avec plusieurs unités phonétiques à la fois. Réciproquement, chaque unité phonétique se matérialise par des indices acoustiques distribués en différents points de ce signal.

Les phénomènes de coarticulation ont donné lieu à de nombreuses études dans le domaine de la perception, la question étant de savoir comment s'opère le décodage d'un signal marqué par la superposition partielle ou totale des indices associés à différentes unités phonétiques. Il est clairement établi que ces effets de superposition ont une influence sur la perception de la parole. On sait ainsi que la perception des fricatives, pour ne prendre que cet exemple, est soumise à l'influence du degré d'arrondissement labial de la voyelle suivante (Mann et Repp, 1980) : une fricative ambiguë, à mi-chemin entre /s/ et /ʃ/, est identifiée différemment selon que la voyelle qui suit est non-arrondie (ex. /ɑ/) ou arrondie (ex. /u/). Loin de compliquer la tâche de l'auditeur, les effets de coarticulation constituent une source d'information mise à profit dans le traitement de la parole, en permettant en particulier que soit anticipée l'émergence dans le signal d'un ou de plusieurs indices acoustiques. On a montré en fait, à de nombreuses reprises, que ces effets sont suffisamment marqués pour donner à un auditeur la possibilité d'identifier correctement des syllabes dont une portion a été supprimée. Les expériences réalisées par Winitz *et al.* (1971), Ostreicher et Sharf (1976), Fowler (1984), entre autres exemples, indiquent que le bruit d'explosion d'une occlusive suffit à identifier la voyelle adjacente avec un taux de réussite supérieur au hasard, lorsque cette voyelle a été supprimée. Réciproquement, un auditeur se montre capable, dans une certaine mesure, de prédire l'identité d'une consonne obstruante ou nasale supprimée à partir de la voyelle adjacente (Ali *et al.*, 1971 ; Ostreicher et Sharf, 1976 ; Pols et Schouten, 1978). Ces travaux, et bien d'autres, ont fait ainsi apparaître que les effets de coarticulation sont exploités par l'auditeur lorsque la principale source d'information sur la séquence à identifier a été soustraite du signal acoustique.

L'influence des effets de coarticulation se montre en outre suffisamment forte pour continuer de se manifester quand sont préservés les principaux indices acoustiques associés à la séquence à identifier. Le plus souvent, cette influence a été établie à partir de stimuli construits par transcollage *(cross-splicing)*. Cette technique consiste par exemple à juxtaposer une consonne (ex. [s]) à une voyelle produite dans le contexte d'une autre consonne (ex. [a] tiré de la séquence [ʃa]). Par suite, les transitions de formant en début de voyelle sont rendues inappropriées au lieu d'articulation de la consonne précédente. Cette rupture de correspondance phonétique *(phonetic mismatch)* a généralement une incidence trop faible sur la

perception pour donner lieu à des erreurs d'identification. Elle peut cependant entraîner un allongement du temps de réaction (ci-après TR), lorsqu'il est demandé au sujet d'identifier le stimulus aussi vite que possible. Cet allongement est interprété comme apportant la preuve que les effets de coarticulation (dans l'exemple qui précède, la transition consonne-voyelle) sont pris en compte dans le traitement de la séquence à l'étude. Les expériences réalisées par Martin et Bunnell sur l'anglais américain (Martin et Bunnell, 1981, 1982) ont ainsi montré que la modification par *cross-splicing* de la première syllabe, dans une séquence de type CVCV, engendrait un allongement du temps demandé pour détecter la seconde voyelle. Selon Martin et Bunnell, ce résultat indique que l'auditeur s'emploie à tirer parti des effets de coarticulation entre voyelles, en cherchant à déterminer de façon anticipée l'identité de V2 à partir des indices qui peuvent lui être rattachés dans V1 (voir également Fowler et Smith, 1986 et Whalen, 1984, pour des expériences construites sur le même modèle).

Malgré la profusion des travaux suscités par les effets de coarticulation dans le domaine perceptif, pendant longtemps on s'est peu interrogé sur le rôle possible de ces effets dans la reconnaissance des mots. Cette absence est sans doute attribuable en partie à la division du travail instituée entre phonétique et psycholinguistique (Frauenfelder, 1992), en vertu de laquelle les phonéticiens se sont pendant longtemps peu intéressés à l'accès au lexique, alors que les psycholinguistes ne prêtaient guère attention pour leur part à la structure détaillée du signal de parole. Mais il est possible aussi que les effets de coarticulation aient été jugés de prime abord trop ténus pour que leur influence puisse se frayer un chemin jusqu'au lexique. Lorsqu'il est par exemple demandé à des sujets d'identifier des voyelles ou des consonnes de synthèse sur un continuum à cheval entre deux catégories phonémiques (perception catégorielle), l'influence du contexte phonétique est souvent confinée aux stimuli les plus ambigus (au centre du continuum). De la même manière, le taux de réussite observé dans l'identification d'une voyelle à partir d'une consonne adjacente se montre soumis à certaines limites (toujours inférieur à 100 %, même dans les cas les plus simples, lorsque les sujets ont à choisir entre deux voyelles seulement ; cf. Fowler, 1984 ; Katz *et al.*, 1991). Dans une tâche de détection de cible, enfin, l'allongement du temps de réponse observé en présence d'un *mismatch* phonétique (engendré par *cross-splicing*) est le plus souvent lui aussi de degré limité, entre 10 ms (Martin et Bunnell, 1981, exp. 1) et 50 ms (Fowler et Smith, 1986) environ.

Les effets de coarticulation exercent donc sur la perception une influence qui, pour être systématique, semble parfois relativement ténue. Cependant, et pour cette raison-là justement, leur possible mise en jeu dans la reconnaissance des mots n'en mérite pas moins d'être étudiée avec le plus grand soin. Pour le phonéticien, qui accorde depuis longtemps une place centrale aux effets de coarticulation, il est sans doute important d'établir jusqu'à quel point leur incidence peut s'étendre dans le traitement de la parole. Pour qui s'intéresse à la reconnaissance des mots en tant que telle,

l'intervention possible des effets de coarticulation dans ce processus soulève une série de questions majeures, relatives aussi bien au « grain » des représentations d'entrée, qu'aux mécanismes mis en œuvre dans l'accès au lexique, ou à la structure des représentations lexicales.

Le plan de cet article est le suivant. En premier lieu, nous passerons en revue un ensemble de résultats expérimentaux récemment obtenus sur le rôle de la coarticulation dans l'identification d'un mot (2). Nous indiquerons ensuite de quelle manière ces résultats sont interprétables par les modèles actuels de la reconnaissance des mots, en prêtant plus particulièrement attention au modèle TRACE et au modèle Cohort (3). Dans un dernier temps (3.3), nous montrerons en quoi les effets de coarticulation nous fournissent des indications nouvelles sur l'architecture générale d'un modèle du traitement de la parole, en présentant une série d'arguments en faveur d'une approche non phonémique.

2. DONNÉES EXPÉRIMENTALES

Les données expérimentales présentées ici ont été regroupées selon la nature de la tâche utilisée et la méthode de construction des stimuli. Nous verrons que cette classification est moins artificielle qu'il n'y paraît, dans la mesure où elle aboutit à rassembler des expériences qui abordent le plus souvent un même problème empirique dans une perspective identique.

2.1. *GATING*

La méthode de présentation de stimuli auditifs dite du dévoilement graduel (*gating*, voir Grosjean, 1980) consiste à présenter un mot-cible par morceaux, ou portes, de durée croissante, les sujets ayant pour tâche de deviner après chaque porte quel est ce mot. Le *gating* offre la possibilité d'étudier la manière dont sont perçus les phénomènes de coarticulation régressive, telle que la présence dans une voyelle d'indices acoustiques associés à la consonne suivante. La technique permet en fait de déterminer avec précision à partir de quel point dans le signal il devient possible au sujet de prédire une unité phonétique à venir. On peut alors tenter de mettre en relation les réponses obtenues avec les changements observés dans la structure acoustique du signal de parole.

Pour plusieurs d'entre eux, les travaux rangés dans cette catégorie ont trait aux phénomènes de nasalisation vocalique. On désigne ainsi le fait qu'une voyelle phonologiquement non nasale est susceptible lorsqu'elle est placée au voisinage d'une consonne nasale de présenter elle-même un certain degré de nasalisation (ce qui veut dire que le voile du palais est abaissé dans la production de cette voyelle au point de laisser entrer l'air en provenance des poumons à l'intérieur de la cavité nasale). Ce phénomène se ren-

contre dans une grande variété de langues, et en particulier en anglais. Une voyelle peut donc fournir à l'auditeur des indices sur le caractère nasal ou non nasal de la consonne subséquente, selon qu'elle est elle-même nasalisée ou pas. Dans une expérience sur l'anglais britannique, Warren et Marslen-Wilson (1987) ont ainsi montré qu'un auditeur est capable de déterminer si un mot monosyllabique se termine par une consonne nasale (ex. « drown », [draʊn]) ou non nasale (ex. « drought », [draʊt]) dès la fin de la voyelle. Ces résultats ont été répliqués par Ohala et Ohala (1995, anglais canadien) et par Kearns et Nguyen (1997, anglais britannique). En revanche, dans une autre expérience réalisée sur le même modèle par Lahiri et Marslen-Wilson (1991), les mots anglais de type CVN utilisés ne se sont pas révélés clairement reconnaissables avant le début de la consonne nasale finale.

Le rôle de la nasalisation vocalique dans la reconnaissance des mots a également été étudié dans certaines langues indo-aryennes telles que le bengali (Lahiri et Marslen-Wilson, 1991), et l'hindi (Ohala et Ohala, 1995), pour lesquelles l'opposition entre voyelles nasales et voyelles non nasales revêt à la différence de l'anglais une valeur distinctive. Des études similaires (Ingram et Mylne, 1994 ; Kearns et Nguyen, 1997) ont été menées en français, langue dont le système vocalique est encore différent, puisqu'à la distinction nasal/non-nasal (ex. « paix »-« pain », « pas »-« pan », etc.) sont associées de multiples différences relatives aussi bien au lieu d'articulation qu'au degré d'ouverture et au degré d'arrondissement labial (voir, par ex., Zerling, 1984), et que les effets de nasalisation dits contextuels (nasalisation d'une voyelle phonologiquement non nasale placée dans le voisinage d'une consonne nasale) sont de degré plus limité que dans les autres langues citées.

La technique du *gating* a également été employée pour étudier le rôle dans l'identification des mots des effets de coarticulation liés au lieu d'articulation, au mode d'articulation (interrompu/continu) et au voisement. Dans l'expérience de Warren et Marslen-Wilson (1987), l'auditeur se voyait ainsi présenter des mots dont la différence résidait dans le lieu d'articulation (ex. « scoop » [skup] *vs* « scoot » [skut]) ou dans le mode d'articulation (ex. « spout » [spaʊt] *vs* « spouse » [spaʊs]) de la consonne finale. Les résultats montrèrent que le lieu d'articulation de la consonne finale commençait à être correctement identifié par l'auditeur 80 ms au moins avant la fin de la voyelle, bien que le pourcentage de réponses correctes relevées à la fin de cette voyelle ne fût pas supérieur à 55 % (contre 30 % de réponses associées au lieu d'articulation opposé). En revanche, la voyelle ne fournissait apparemment à l'auditeur aucun indice sur le caractère continu ou interrompu de la consonne finale, puisque les fricatives ne commençaient à être différenciées des occlusives correspondantes qu'une fois le début de la consonne entendu. Le traitement des indices relatifs au lieu d'articulation dans la reconnaissance des mots est exploré de manière plus détaillée dans Warren et Marslen-Wilson (1988) et Marslen-Wilson et Warren (1994). Warren et Marslen-Wilson (1988) ont également établi que le caractère voisé ou non voisé d'une occlusive finale est lui aussi

susceptible d'être déterminé de manière anticipée, à partir de la durée de la voyelle précédente, dans une tâche de reconnaissance de mots.

En résumé, les expériences que nous venons de citer ont montré que les effets de coarticulation jouent un rôle important dans la reconnaissance des mots. On constate que ces effets permettent qu'un mot présenté isolément soit dans certains cas reconnu avant le point d'unicité, c'est-à-dire le point à partir duquel le mot devient unique dans le lexique. Peut-être serait-il d'ailleurs nécessaire de faire entrer les effets de coarticulation en ligne de compte dans la définition du point d'unicité, en substituant à la représentation phonémique abstraite à partir de laquelle la position de ce point est généralement établie, une représentation phonétique plus détaillée.

Néanmoins, les travaux évoqués dans la présente section sont soumis à certaines limites, pour partie inhérentes à la méthode de présentation des stimuli utilisée. On a parfois mis en avant le fait que le *gating* pouvait donner lieu à des stratégies de réponse spécifiques, dans la mesure où l'auditeur cherche vraisemblablement à associer chaque stimulus avec toutes les représentations lexicales susceptibles de lui correspondre au point où le signal prend fin. En d'autres termes, chaque porte est interprétée dans la mesure du possible comme un mot entier. Par ailleurs, notons que les stimuli utilisés donnent rarement lieu à des analyses acoustiques. Il est pourtant permis de penser que c'est en examinant la structure acoustique détaillée de ces stimuli qu'il serait possible d'expliquer les divergences parfois observées entre différentes études réalisées sur des phénomènes coarticulatoires de même nature (ex. nasalisation de voyelle).

2.2. TÂCHE DE DÉCISION LEXICALE

La tâche de décision lexicale permet d'étudier l'influence de la coarticulation dans la reconnaissance des mots telle que cette influence se manifeste « en ligne », lorsqu'il est demandé au sujet de déterminer aussi vite que possible si le stimulus auditif présenté est un mot ou un non-mot. Dans toutes les expériences citées, une comparaison est établie entre les temps de réaction (ci-après TR) obtenus pour un groupe de stimuli naturels, et les TR obtenus pour un autre groupe de stimuli construits à partir des premiers par *cross-splicing*, de la manière décrite plus haut. L'avantage de cette méthode de mesure est qu'elle se montre sensible à des effets de coarticulation très fins, dont l'influence peut passer inaperçue dans une tâche d'identification lexicale simple par exemple.

Dans une expérience réalisée sur ce modèle, Streeter et Nigro (1979) ont étudié le rôle des transitions VC dans la reconnaissance de mots dissyllabiques (anglais américain). Dans les stimuli construits par *cross-splicing*, les transitions VC et CV fournissaient sur la consonne intervocalique des indices divergents. Ce *mismatch* phonétique était relatif soit au voisement de la consonne (ex. « sta(ple) » + « (sta)ble »), soit au lieu d'articulation (ex. « fa(ded) » + « (fa)ble »), soit encore au mode d'articulation (ex.

« trai(tor) » + « (tra)ces »). Les résultats firent apparaître qu'un *mismatch* phonétique entraînait en moyenne un allongement du temps de réaction d'environ 75 ms pour les mots, les stimuli originaux servant de base de comparaison. En revanche, aucune variation significative du TR ne fut observée entre les stimuli originaux et les stimuli cross-splicés, en ce qui concerne les non-mots. Il est intéressant de noter que dans un test d'intelligibilité réalisé en parallèle, le *cross-splicing* s'était révélé sans effet sur le taux de réponses correctes, indépendamment du statut lexical du stimulus.

La même méthode fut employée par Whalen (1991) dans une étude destinée à déterminer l'importance des transitions entre fricative et voyelle dans une tâche de décision lexicale portant sur des mots monosyllabiques en anglais américain. Là également, les stimuli cross-splicés donnèrent lieu à des réponses plus tardives que les stimuli originaux. Cependant, cette différence était en moyenne extrêmement réduite (+ 16 ms pour les mots, + 5 ms pour les non-mots), ce qui s'explique peut-être par le fait que les transitions de formant sont généralement considérées comme ayant un rôle mineur dans l'identification du lieu d'articulation pour les fricatives utilisées dans cette expérience (sibilantes non voisées ; cf. Harris, 1958). Dans une étude plus récente menée sur l'anglais britannique, Marslen-Wilson et Warren (1994) ont également observé que les transitions voyelle-consonne ne semblaient pas avoir d'influence sur la vitesse de décision lexicale, lorsque la consonne postvocalique était une fricative non voisée ou une occlusive non voisée. Pour les stimuli se terminant par une occlusive voisée (ex. « job », « jog »), en revanche, les TR se montrèrent beaucoup plus sensibles aux effets de *mismatch* phonétique, un retard moyen de + 122 ms pour les mots et de + 52 ms pour les non-mots étant enregistré lorsque la transition voyelle-consonne et le bruit d'explosion étaient associés à des lieux d'articulation différents (ex. « jo(g) » + « (jo)b »), par comparaison avec les stimuli de contrôle (« jo(b) » + « (jo)b »).

Andruski *et al.* (1994) ont examiné l'effet sur la reconnaissance des mots de petites différences dites « subphonétiques », telles qu'il en existe parmi les sons rattachés à un même trait distinctif. L'étude a plus spécifiquement porté sur les variations aléatoires dans la durée du VOT *(voice onset time)* présentées par les occlusives non voisées en début de mot en anglais[1]. L'expérience avait pour but de déterminer si des modifications artificielles apportées au VOT (suffisamment petites pour l'occlusive reste perçue comme non voisée) pouvaient avoir une influence sur le niveau d'activation du mot porteur dans le lexique. Dans cette expérience, les sujets avaient à effectuer une tâche de décision lexicale sur des mots et des non-mots présentés auditivement. Dans la condition test, chaque item était précédé par un mot qui lui était relié sémantiquement, et qui débutait par une occlusive non voisée. Les résultats ont fait apparaître que les effets

1. Notons qu'il ne s'agit pas là d'un effet de coarticulation au sens strict du terme.

d'amorçage sémantiques étaient plus réduits lorsque l'occlusive initiale de l'amorce avait été modifiée par raccourcissement du VOT.

Plus récemment encore, Hawkins et Nguyen (1999) se sont intéressés au rôle possible dans la reconnaissance des mots de certains indices acoustiques très fins pouvant être associés au voisement d'une occlusive finale, en anglais britannique. Comme on le sait, la durée d'une voyelle est plus longue lorsque cette voyelle se trouve placée devant une occlusive voisée, plutôt que devant une occlusive non voisée. Selon de récents travaux cependant (van Santen *et al.*, 1992), ces variations de durée ne se produisent pas de manière uniforme sur toute la longueur de la voyelle : elles se montrent plus importantes en début de voyelle qu'en fin de voyelle, et sont également susceptibles de s'étendre à la consonne précédente lorsque cette consonne est une sonante. Il a été établi en particulier qu'un /l/ en position prévocalique est légèrement plus long dans une syllabe se terminant par une occlusive voisée (ex. « led ») plutôt qu'une occlusive non voisée (« let »). Nous avons montré en outre que de telles variations de durée étaient associées à des variations spectrales (/l/ plus sombre lorsque l'occlusive est voisée plutôt que non voisée ; Nguyen et Hawkins, 1998). Hawkins et Nguyen ont entrepris d'évaluer l'importance de ces petites variations acoustiques dans la reconnaissance des mots, en examinant si un auditeur serait plus lent à déterminer le statut lexical d'une séquence monosyllabique, lorsque le /l/ initial n'était pas susceptible de fournir à cet auditeur des indices sur l'occlusive finale (ex. « l(oat) » + « (l)oad », par opposition à la séquence originale, « load »). Les résultats firent apparaître que la différence entre les TR pour les stimuli cross-splicés et les TR pour les stimuli originaux n'était pas significative. Une analyse *posthoc* révéla cependant que pour les mots non voisés, cette différence de TR augmentait lorsque l'écart observé sur le plan acoustique entre le /l/ d'origine et le /l/ entendu (durée/fréquence du deuxième formant) était plus grand. L'expérience semble ainsi montrer que des indices acoustiques ténus et associés à un composant éloigné dans une syllabe peuvent, dans certaines conditions du moins, exercer une influence dans l'identification des mots.

En résumé, les travaux que nous venons de passer en revue donnent à nouveau à penser que l'auditeur prête attention aux phénomènes de coarticulation dans la reconnaissance des mots. La tâche de décision lexicale présente l'intérêt de soumettre le sujet à des contraintes temporelles analogues à celles qui se rencontrent dans le traitement de la parole naturelle. En situation de communication ordinaire, l'auditeur est souvent amené à identifier plusieurs mots par seconde. Dans une tâche de décision lexicale, il est demandé au sujet de fournir une réponse en ligne, c'est-à-dire séparée du stimulus par un intervalle temporel aussi bref que possible. Placé dans une telle situation, on peut supposer que l'auditeur cherche à faire appel à toutes les données sensorielles dont il dispose pour catégoriser le stimulus. Lorsque de telles contraintes temporelles s'imposent à l'auditeur, des effets de coarticulation extrêmement fins se révèlent avoir une incidence sur la reconnaissance des mots.

2 . 3. AMORÇAGE TRANSMODAL

Dans cette section sont rassemblés de récents travaux ayant eu pour but d'examiner la manière dont sont traités les phénomènes dits d'assimilation[1], dans la reconnaissance des mots. Ces travaux ont porté plus particulièrement sur les variations présentées par les occlusives alvéolaires en fin de mot sous l'influence de la consonne suivante, en anglais. Dans la séquence *that man*, par exemple, le lieu d'articulation de la consonne /t/ peut en parole continue être assimilé à celui de l'occlusive bilabiale subséquente (la forme de surface résultante étant alors [ðæp mæn]). L'intérêt, pour ce qui concerne la reconnaissance des mots, de ce mécanisme assimilatoire, est qu'il se produit à cheval entre deux mots. On s'est ainsi demandé si l'auditeur était sensible à des modifications ayant lieu à la fin du premier mot, c'est-à-dire au-delà souvent du point d'unicité de celui-ci, et l'on a également cherché à savoir s'il était nécessaire à l'auditeur d'avoir entendu le début du second mot pour identifier correctement le premier. Ces questions ont des implications théoriques importantes, puisqu'elles peuvent laisser supposer que l'identification des mots ne s'opère pas de manière strictement linéaire (un mot après l'autre), et peut donner lieu à des retours en arrière.

Une expérience réalisée par Marslen-Wilson et Gaskell (1992) tend à montrer que la reconnaissance des mots peut effectivement être perturbée par un *mismatch* phonologique localisé à la fin d'un mot et postérieur au point d'unicité. Dans cette expérience, les sujets avaient à effectuer une décision lexicale sur des mots-cible présentés visuellement, chaque mot-cible étant précédé par une amorce auditive présentant ou non une relation sémantique avec ce mot (*cross-modal priming*, ou amorçage transmodal). Les résultats firent apparaître qu'un mot-cible était traité plus rapidement lorsque l'amorce lui était associée sémantiquement (ex. amorce = « grotesque », cible = « ugly »), une différence moyenne de – 30 ms étant enregistrée par rapport à la condition de contrôle (pas de lien sémantique amorce-cible). Cependant, cet effet d'amorçage sémantique disparaissait en présence d'un *mismatch* phonologique à la fin de l'amorce (ex. « grotest » au lieu de « grotesque »).

En revanche, dans un travail faisant appel à une méthode un peu différente (tâche de répétition avec amorçage transmodal), Nix *et al.* (1993) ont

1. On a longtemps établi une distinction de nature entre coarticulation et assimilation, en attribuant aux phénomènes de coarticulation un caractère graduel et aux phénomènes d'assimilation un caractère discret. Cette distinction est considérée aujourd'hui comme étant assez artificielle, de récentes études menées sur les mouvements articulatoires dans la production de la parole ayant montré que les phénomènes d'assimilation sont eux aussi susceptibles de présenter des variations de degré (voir, ainsi, la notion de geste alvéolaire résiduel, Byrd, 1992 ; Nolan, 1992).

établi qu'un *mismatch* phonologique en fin de mot ne semble pas rendre ce mot plus difficile à reconnaître lorsque celui-ci se présente dans un contexte phonologique approprié. Chaque amorce figurait à l'intérieur d'une phrase, devant un autre mot (ex. « wicked prank »). Cette amorce pouvait ou non être marquée par un *mismatch* phonologique relatif à la consonne finale (ex. « wickib », par opposition à « wicked »), lequel *mismatch* pouvait ou non être interprété comme un effet d'assimilation induit par la consonne suivante (ex. « wickib prank », par opposition à « wickib game »). Le mot-cible était présenté sur un écran immédiatement après l'amorce. Là aussi, un effet d'amorçage fut mis en évidence, les temps de réaction étant en moyenne plus courts lorsque l'amorce et la cible étaient associées l'une avec l'autre que lorsqu'elles ne l'étaient pas. En outre, la présence d'un *mismatch* phonologique à la fin de l'amorce ne donnait pas lieu à une diminution de l'effet d'amorçage, lorsque ce *mismatch* pouvait être attribué par l'auditeur à l'influence de la consonne subséquente (ex. « wickib prank »). Selon Nix *et al.*, ce résultat donne à penser que des règles d'inférence phonologique, visant à reconstruire les segments sujets à des effets d'assimilation, sont mises en application dans la reconnaissance des mots. Ces règles stipuleraient ainsi que l'occlusive bilabiale finale de *wickib* peut s'interpréter comme une alvéolaire dont le lieu d'articulation a été assimilé à celui de la consonne suivante dans la séquence *wickib prank*. L'expérience présentée par Nix *et al.* fait l'objet d'une description plus détaillée dans Gaskell et Marslen-Wilson (1993, 1996). L'influence sur l'accès au lexique des effets d'assimilation opérant sur les occlusives alvéolaires en fin de mot est également traitée dans Shockey et Watkins (1995) et Sotillo *et al.* (1995), pour l'anglais, et dans van Heuven et Jongenburger (1993) pour le hollandais.

2.4. AUTRES ÉTUDES

Mentionnons pour conclure cette partie expérimentale différentes études dont l'objectif ne fut pas à proprement parler d'examiner le rôle de la coarticulation dans la reconnaissance des mots, mais plutôt de déterminer la manière dont se combinent effets de coarticulation et effets lexicaux dans le traitement de la parole. Par effet lexical nous désignons le fait que le lexique exerce une influence sur la façon dont les sons de la parole sont catégorisés, dans une tâche d'identification de phonème par exemple. La célèbre expérience de Ganong (1980) a ainsi montré que lorsque des sujets ont à identifier une occlusive sur un continuum entre un mot (ex. « dash ») et un non-mot « tash », ils optent plus fréquemment pour la réponse formant un mot avec la séquence porteuse (dans l'exemple cité, « d »). Le lexique est vu ici comme l'une des sources d'information susceptibles d'entrer en jeu dans une tâche de cette nature. De récentes expériences ont eu pour objectif d'examiner les interactions possibles, au sens large du terme, entre effets lexicaux et effets de coarticulation, dans le but d'en infé-

rer des indications plus générales sur l'architecture du système de traitement.

Le travail d'Elman et McClelland (1988) avait pour principal objectif de montrer qu'un transfert d'information s'opère de haut en bas, depuis le lexique vers le niveau phonémique, dans le traitement de la parole. Dans cette perspective, Elman et McClelland ont cherché à établir si les informations contenues dans le lexique sont en mesure d'induire des effets de contexte latéraux entre deux phonèmes adjacents. Dans un tel cas, cela montrerait, selon eux, que le lexique a une incidence directe sur l'identification des phonèmes, comme cela est postulé dans le modèle TRACE (McClelland et Elman, 1986). Elman et McClelland ont construit leur expérience à partir d'un effet de contexte observé par Mann et Repp (1981) dans les séquences fricative + occlusive. Il a été montré qu'une occlusive à mi-chemin entre [t] et [k] est catégorisée différemment selon que la fricative précédente est alvéolaire ([s]) ou postalvéolaire ([ʃ]) : cette occlusive est identifiée plus fréquemment comme une vélaire ([k]) lorsqu'elle est précédée par [s] plutôt que [ʃ]. L'expérience menée par Elman et McClelland visait à déterminer si cet effet continue d'avoir lieu en présence d'une fricative elle-même ambiguë, mais dont le lexique permette de rétablir l'identité. Chaque stimulus se présentait sous la forme d'une séquence de deux mots dont le premier se terminait par une fricative ambiguë (ex. « ChristmaS », « fooliS », S représentant une fricative à mi-chemin entre /s/ et /ʃ/), et dont le second commençait par une occlusive sur un continuum entre /t/ et /k/ (ex. « ?ape »). Les sujets avaient pour tâche d'identifier cette occlusive. Les réponses observées présentaient les variations attendues en fonction du mot précédent : l'occlusive était plus fréquemment associée à [k] dans le contexte de « ChristmaS » plutôt que « fooliS ». Partant de l'hypothèse que ces effets de contexte entre fricatives et occlusives sont de nature perceptive, Elman et McClelland en ont conclu que le lexique exerçait une influence sur la manière même dont la fricative était perçue, autrement dit, que l'identification des phonèmes faisait bien intervenir des processus de traitement de type *top-down*.

Serniclaes *et al.* (1995) ont examiné la relation entre effets lexicaux et effets du contexte phonétique sous un angle plus méthodologique, en cherchant à savoir si le contexte phonétique pouvait constituer une source d'artefact expliquant ces fortes disparités observées dans la taille des effets lexicaux d'une étude à l'autre (Pitt et Samuel, 1993). Ces auteurs soulignent que, dans une expérience visant à étudier l'influence du lexique sur la perception d'un trait distinctif (ex. voisement), la structure phonétique de la séquence porteuse utilisée demande à être contrôlée avec le plus grand soin puisqu'elle est elle aussi susceptible d'avoir un effet sur les réponses obtenues. Serniclaes *et al.* se sont penchés plus particulièrement sur l'opposition entre occlusives voisées et occlusives non voisées en position initiale. Selon eux, le principal paramètre phonétique associé à cette opposition *(voice onset time)* présente des variations selon la nature de la consonne finale, dans une syllabe fermée. Toutes choses égales d'ailleurs, le

VOT serait, par exemple, plus long dans la séquence « type » que dans la séquence « tice », pour des raisons de nature aérodynamique. En réexaminant la littérature sur les effets lexicaux à la lumière de cette hypothèse, Serniclaes *et al.* aboutissent à la conclusion que la variabilité apparente de ces effets peut bien dans certains cas s'expliquer par des différences relatives à la structure phonétique des séquences porteuses.

Nguyen (1995) a cherché à déterminer la contribution respective du contexte phonétique et du lexique dans l'identification des fricatives en français, lorsque ces deux sources d'information fournissent des indices divergents sur le phonème-cible. Les fricatives à identifier prenaient place sur un continuum /s/-/ʃ/, et elles se présentaient dans des séquences dissyllabiques choisies de manière à croiser les effets de contexte phonétiques avec les effets lexicaux. On sait ainsi que la distinction entre /s/ et /ʃ/ s'opère différemment selon que la voyelle subséquente est arrondie ou non arrondie : une fricative ambiguë, à mi-chemin entre /s/ et /ʃ/, est plus souvent identifiée comme étant un /s/ devant une voyelle arrondie plutôt que devant une voyelle non-arrondie (Mann et Repp, 1980). Les séquences dissyllabiques employées dans cette expérience étaient ainsi telles que le contexte phonétique et le lexique pouvaient pousser le sujet soit à produire la même réponse (ex. /s/ lorsque la fricative précédait la séquence « apin », « sapin » étant un mot par opposition à « chapin » et /a/ une voyelle non arrondie), soit à produire des réponses opposées (quand la fricative était accolée à la séquence « oucroute » par ex., le lexique devait favoriser la réponse /ʃ/ dans la mesure où « choucroute » est un mot par opposition à « soucroute », alors que /u/ devait au contraire favoriser la réponse /s/). Le continuum [s]-[ʃ] se composait de 11 stimuli, lesquels ont été présentés au sein de 16 séquences porteuses différentes. Pour la moitié d'entre elles, /s/ formait un mot en combinaison avec la séquence porteuse et [ʃ] un non-mot (ex. « soulier »-« choulier »). Pour l'autre moitié, /ʃ/ formait un mot en combinaison avec la séquence porteuse, et [s] un non-mot (ex. « sapeau »-« chapeau »). Le statut lexical de la séquence porteuse a été croisé avec l'identité de la voyelle faisant suite à la fricative (/a/, /u/). Les sujets avaient pour tâche d'identifier la fricative (« s » ou « ch »). La figure 1 illustre les principaux résultats obtenus.

Les pourcentages moyens de réponses « s » sont représentés selon la position de la fricative sur le continuum [s]-[ʃ] ([s] : stimulus 1, [ʃ] : stimulus 11). Comme cela était attendu, ce pourcentage diminue entre le premier et le dernier stimulus sur le continuum. À gauche, les pourcentages sont décomposés en fonction de la voyelle suivante (/a/ ou /u/), de façon à mettre en évidence l'effet de la voyelle sur la manière dont les fricatives sont identifiées. Conformément aux résultats de Mann et Repp (1980), on voit que les sujets tendaient à identifier les fricatives plus souvent comme /s/ lorsque celles-ci étaient placées dans le voisinage d'une voyelle arrondie (/u/). Les courbes de droite font apparaître l'effet du lexique sur les réponses des sujets (○ : [s] forme un mot avec la séquence porteuse et [ʃ] un non-mot, ex.

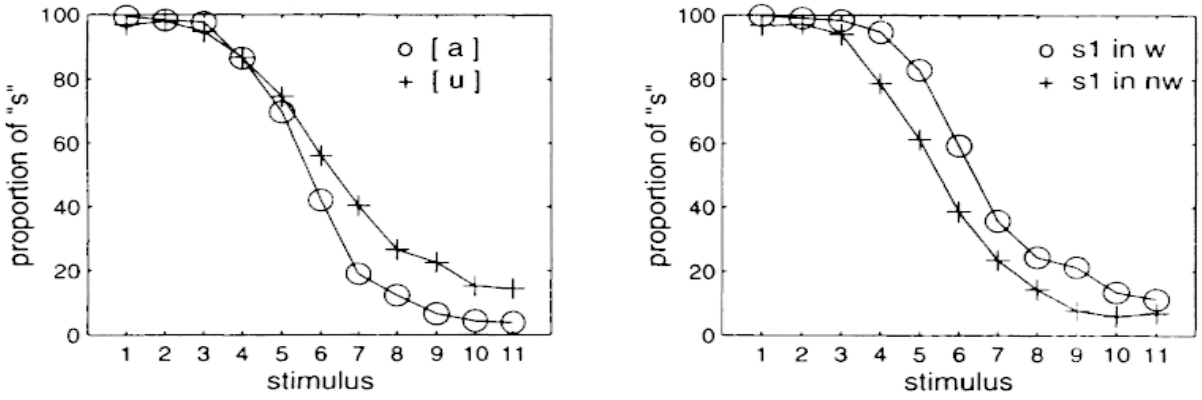

Fig. 1. — Pourcentages moyens de réponses « s » pour chacune des 11 fricatives sur le continuum [s]-[ʃ]. À gauche : pourcentages décomposés en fonction de la voyelle suivante. À droite : pourcentages décomposés selon le statut lexical de la séquence porteuse. Tiré de Nguyen (1995)

Average percentages of « s » responses, for each of the 11 fricatives on the* [s]-[ʃ] *continuum. Left panel : percentages broken down depending on the following vowel. Right panel : percentages broken down depending on the lexical status of the carrier string. From Nguyen (1995)

« soulier »-« choulier » ; + : [s] forme un non-mot avec la séquence porteuse et [ʃ] un mot, ex. « sapeau »-« chapeau »). En accord avec nos attentes, les fricatives tendaient à être plus souvent identifiées comme /s/ lorsque /s/ formait un mot en combinaison avec la séquence porteuse.

Les résultats firent apparaître que l'auditeur se montre sensible dans une telle situation à la fois à l'influence du lexique et à celle du contexte phonétique. Il est intéressant de remarquer que le lexique ne semble donc pas primer sur le contexte phonétique lorsqu'ils sont l'un et l'autre susceptibles d'avoir une influence sur les réponses des sujets. Les effets de contexte phonétiques étudiés dans cette expérience étaient suffisamment robustes pour continuer de s'observer alors que le lexique était amené à entrer en jeu. Notons toutefois que l'attention des auditeurs était probablement focalisée sur la structure phonétique des stimuli, en raison de la nature de la tâche utilisée (catégorisation phonémique). Des résultats analogues ont été obtenus par Nguyen, Wrench, Gibbon et Hardcastle (1998) dans une expérience similaire sur l'identification des occlusives en anglais.

En résumé, les expériences mentionnées dans cette section ont permis d'établir que les effets lexicaux et les effets de coarticulation donnent lieu à un processus d'intégration perceptive, dans une tâche d'identification de phonème. En d'autres termes, la réponse de l'auditeur est déterminée à partir d'une combinaison entre ces deux sources d'information. Ces résultats nous fournissent une nouvelle preuve de l'importance des effets de coarticulation dans le traitement de la parole.

3. COARTICULATION ET MODÈLES DE LA RECONNAISSANCE DES MOTS

Les données expérimentales dont nous disposons aujourd'hui montrent donc que les effets de coarticulation jouent bien un rôle dans la reconnaissance des mots. Ces résultats revêtent une importance sur le plan méthodologique. Dans toute expérience sur la reconnaissance des mots parlés, il est clair que la structure phonétique détaillée des stimuli utilisés doit être établie avec soin. Aussi ténus soient-ils, les effets de coarticulation sont utilisables par l'auditeur pour anticiper la manière dont un mot se termine. Ces effets peuvent permettre qu'un mot soit reconnu plus tôt qu'il est possible de le prédire à partir d'une représentation phonémique abstraite de ce mot. Les phénomènes de coarticulation demandent à être considérés avec la plus grande attention à chaque fois que le point d'unicité ou les voisins phonétiques d'un mot sont à déterminer par l'expérimentateur.

Les résultats expérimentaux précédemment exposés ont en outre, et surtout, d'importantes implications en ce qui concerne les modèles actuels de la reconnaissance des mots, et c'est à décrire ces implications que nous allons à présent nous attacher. Le lecteur est renvoyé à Frauenfelder (1991) pour une introduction générale aux modèles abordés ci-après.

3.1. COARTICULATION ET REPRÉSENTATIONS DE BASE DANS LA RECONNAISSANCE DES MOTS

La discussion qui suit sera fondée pour l'essentiel sur un exemple tiré de Kearns et Nguyen (1997). Cette expérience a fait apparaître qu'en anglais, un auditeur se montre capable d'établir une différence entre deux mots monosyllabiques tels que « said » (/sed/) et « send » (/send/) dès la fin de la voyelle, lorsque ces mots lui sont présentés selon la technique du *gating*. En d'autres termes, l'auditeur recueille dans la voyelle des indices acoustiques lui permettant de déterminer le caractère oral ou nasal de la consonne postvocalique. La question est de savoir de quelle manière un modèle de la reconnaissance des mots peut rendre compte de ce phénomène.

Un premier élément de réponse se rapporte à la représentation à partir de laquelle il est postulé que débute l'identification des mots dans le modèle. Nous appellerons cette représentation la représentation de base[1].

1. Notons que cette représentation est à différencier de la représentation d'entrée, si l'on utilise ce dernier terme pour désigner la forme sous laquelle le signal de parole est converti par le système auditif (Frauenfelder et Nguyen, 2000). La plupart des modèles de la reconnaissance des mots se donnent pour point de départ une représentation du signal de parole déjà fort éloignée de cette représentation d'entrée.

Cette représentation de base doit être nécessairement une représentation détaillée reflétant tous les contrastes phonétiques entre « said » et « send », pour autant que ces contrastes soient perceptibles par l'auditeur.

Dans les faits, la plupart des modèles actuels présentent la reconnaissance des mots comme un processus s'appliquant à une représentation abstraite du signal de parole. C'est le cas du modèle TRACE II (McClelland et Elman, 1986), par exemple, dans lequel le signal est représenté sous la forme d'un ensemble de traits qu'il est pour certains d'entre eux difficile d'associer à des corrélats acoustiques bien définis (ex. trait de voisement). Cette remarque s'applique également aux modèles du type réseau de neurones récurrents récemment proposés par Shillcock *et al.* (1993), Gupta et Mozer (1993), ou Gaskell, Hare et Marslen-Wilson (1995). La représentation de base choisie est parfois d'ailleurs suffisamment abstraite pour se montrer appropriée à l'écriture aussi bien qu'à la parole (Norris, 1992). Explicitement désigné sous le terme de *mock speech* (parole factice), ce type de représentation s'est sans doute imposé pour des raisons d'ordre pratique : on s'attache aujourd'hui à simuler la reconnaissance des mots en utilisant des lexiques de taille réaliste, avec un renforcement résultant des contraintes en matière de ressources computationnelles que des représentations de base aussi parcimonieuses que possible peuvent peut-être contrebalancer. Il n'en reste pas moins que le *mock speech* utilisé dans de nombreux modèles de simulation sur la reconnaissance des mots parlés prive ces modèles de pertinence sur le plan phonétique.

Dans l'exemple que nous avons choisi, la représentation de base devrait nécessairement spécifier que la voyelle est (phonétiquement) non nasale dans « said » par opposition à « send ». Cette représentation devrait également contenir des informations relatives à la durée de la nasalité, de façon à refléter les différences observées entre « send » (voyelle nasalisée dans sa deuxième partie) et « sent », par exemple (voyelle entièrement nasalisée ; cf. Local, 1992). En bref, face à l'approche minimaliste prévalant pour ce qui touche aux représentations de base, les expériences les plus récentes sur la perception des effets de coarticulation vont au contraire dans le sens d'une approche maximaliste, fondée sur l'idée que l'identification des mots s'opère à partir d'une représentation phonétique hautement détaillée. Dans la section suivante, nous tentons de déterminer selon quelles voies les effets de coarticulation peuvent mis à profit par l'auditeur dans la reconnaissance des mots.

3 . 2. COARTICULATION ET ACCÈS AU LEXIQUE

Parmi les modèles actuels de la reconnaissance des mots, nous nous limiterons ici à examiner le modèle TRACE (McClelland et Elman, 1986) et le modèle Cohort dans sa version la plus récente (Marslen-Wilson et Warren, 1994). Ces deux modèles occupent, comme on le sait, une place primordiale dans les études sur la reconnaissance des mots (Frauenfelder, 1991).

L'intérêt d'établir une comparaison entre ces deux modèles tient également à ce qu'ils rendent compte des effets de coarticulation de manière extrêmement différente, TRACE étant un modèle phonémique, Cohort un modèle non phonémique.

3.2.1. Le modèle TRACE

Rappelons que TRACE est un modèle de la reconnaissance des mots de type connexionniste, et qu'il se compose d'un grand nombre d'unités de traitement réparties sur trois niveaux séparés : le niveau des traits, le niveau des phonèmes et celui des mots. Ces unités s'apparentent en fait à des détecteurs de trait, de phonème ou de mot. Chaque unité se caractérise par un certain niveau d'activation évoluant dans le temps selon les informations qui lui parviennent sur le mot à reconnaître. Des connexions facilitatrices s'établissent verticalement entre niveaux de traitement adjacents, de bas en haut (trait-phonème et phonème-mot) et de haut en bas (mot-phonème). Des connexions inhibitrices sont établies latéralement entre unités de même niveau. Le système de traitement est mis en route lorsque le signal de parole vient activer la couche des traits. L'activation se propage alors dans le réseau de bas en haut, de la couche des traits jusqu'à celle des mots, et de haut en bas, de la couche des mots vers celle des phonèmes. La mise en relation entre signal de parole et lexique s'accomplit de manière progressive, le système étant placé sous le contrôle d'une horloge interne régissant la vitesse avec laquelle l'information se propage à l'intérieur de chaque couche et d'une couche à l'autre. À ce découpage horizontal en couches vient s'ajouter une série de subdivisions verticales, chaque couche étant constituée d'un certain nombre de colonnes juxtaposées contenant chacune un ensemble complet de détecteurs. Chaque colonne de détecteurs a pour fonction de traiter une section déterminée du signal de parole (dont la durée est de 5 ms pour les détecteurs de trait, et de 15 ms pour les détecteurs de phonème). Cette organisation verticale permet de simuler la manière dont nous percevons un signal temporel tel que le signal de parole en spatialisant le temps, c'est-à-dire en découpant le signal en une suite de morceaux et en dupliquant le réseau autant de fois que l'on a de morceaux (bien évidemment, cette solution n'est praticable que pour des signaux de durée limitée).

Deux propriétés rendent le réseau sensible aux effets de coarticulation. En premier lieu, la fenêtre temporelle à l'intérieur de laquelle une colonne de phonèmes est activable par le signal de parole empiète partiellement sur la fenêtre réservée à la colonne de phonèmes adjacente. Le segment de parole commun à ces deux fenêtres temporelles sera donc simultanément mis en relation avec les deux colonnes de phonèmes. Si le mot « send » est présenté au réseau, par exemple, cela signifie que la partie finale nasalisée du noyau vocalique contribuera à activer à la fois le détecteur associé à la voyelle /e/, et le détecteur associé à la consonne /n/ dans la colonne adjacente. Ce recouvrement temporel entre fenêtres de traitement fournit une

explication simple au fait que le mot « send » puisse être reconnu dès la fin de la voyelle lorsqu'il est dévoilé graduellement à l'auditeur.

En second lieu, des connexions établies entre colonnes permettent aux détecteurs de phonème appartenant à une colonne de venir moduler les liens traits-phonèmes dans la colonne adjacente (Elman et McClelland, 1988). Par suite, la mise en relation entre segment de parole et phonème s'opère sous l'influence des phonèmes voisins. C'est par ce mécanisme qu'il est possible dans TRACE de simuler le fait qu'un même son puisse être catégorisé différemment selon le contexte phonétique dans lequel il se présente. Dans une séquence fricative + voyelle, par exemple, on peut supposer que la façon dont les traits relatifs à la consonne initiale seront interprétés dans le modèle (mis en correspondance avec les phonèmes /s/, /ʃ/, etc.) dépendra partiellement du caractère arrondi ou non-arrondi de la voyelle suivante (voir section 2.4).

La mise en place de connexions entre phonèmes adjacents obéit à une hypothèse classique en vertu de laquelle chaque phonème est identifié d'une manière dépendante du contexte, c'est-à-dire en fonction des phonèmes qui l'entourent. Cette hypothèse repose elle-même sur l'idée que les phonèmes exercent une influence l'un sur l'autre dans la chaîne parlée, autrement dit que chaque phonème est produit de manière différente selon les phonèmes adjacents (« we can define coarticulation as the influence of one speech segment upon another », Daniloff et Hammarberg, 1973). Dans la production de la syllabe « sou » (/su/), par exemple, on supposera que la fricative /s/ est produite avec les lèvres arrondies de manière à assurer une transition plus facile avec la voyelle /u/, elle-même arrondie. Il sera postulé ainsi que la fricative /s/ est modifiée, sur le plan articulatoire et acoustique, au contact de la voyelle subséquente. Dans le domaine perceptif, on supposera que l'auditeur a connaissance de ces effets d'arrondissement contextuels, et prend la voyelle en ligne de compte dans l'identification de la fricative (Mann et Repp, 1980).

Le chevauchement des fenêtres de traitement constitue une hypothèse moins classique dont on a peut-être sous-estimé la portée théorique. En substance, cette hypothèse revient à considérer que le signal de parole se décompose sous la forme d'une séquence de segments partiellement superposés sur l'axe temporel (Fowler, 1980 ; Browman et Goldstein, 1986). Dans le mot « send », par exemple, il sera supposé que la consonne /n/ débute « à l'intérieur » de la voyelle /e/, c'est-à-dire alors que celle-ci n'a pas encore pris fin. Aussi simple soit-elle, cette hypothèse a deux conséquences fondamentales. En premier lieu, elle conduit à abolir la distinction entre phonème « influençant » et phonème « influencé » à laquelle il est fait référence ci-dessus. Il est inapproprié dans un tel cadre de considérer la voyelle comme étant soumise à une modification (nasalisation, en l'occurrence) induite par la consonne suivante. On désignera plutôt ici par coarticulation le fait que les mouvements articulatoires mis en œuvre dans la production de la voyelle et de la consonne sont entrelacés dans le temps, parce qu'ils sont dans une certaine mesure accomplis en parallèle. En

second lieu, la coarticulation devient un phénomène non directionnel. Il n'y a plus lieu dans cette hypothèse de dire que la consonne /n/ est « anticipée » pendant la production de la voyelle précédente, rien ne nous empêchant d'affirmer que c'est au contraire la voyelle qui vient empiéter sur le domaine de la consonne suivante. Sur le plan de la perception, on supposera que l'auditeur perçoit non pas une voyelle nasalisée, mais une voyelle non nasale à l'intérieur de laquelle émerge progressivement la consonne nasale qui suit. Cette hypothèse présente une relation directe avec le modèle d'analyse vectorielle perceptive de Fowler (1984). Elle est en opposition avec l'idée selon laquelle l'identification des phonèmes est un processus de type *context-dependent*. La manière dont les effets de coarticulation sont perçus par l'auditeur est donc simulée dans TRACE par l'entremise de deux mécanismes profondément différents l'un de l'autre.

En accord avec les résultats expérimentaux que nous avons présentés, TRACE donne à supposer que tous les contrastes phonétiques observables entre mots peuvent être mis à profit dans la reconnaissance des mots, pour autant du moins que ces contrastes soient introduits dans la représentation de base. Cela tient d'abord au fait que l'information circule à l'intérieur du réseau sous la forme d'un ensemble de paramètres *continus* (niveaux d'activation des différentes unités de traitement). De petits détails dans la structure acoustique du signal se traduiront par des variations quantitatives dans le niveau d'activation des détecteurs de phonème. Ces variations auront à leur tour un effet sur le niveau d'activation des détecteurs de mots. Le processus de reconnaissance des mots ne serait pas directement sensible à ces petits détails s'il était supposé que les phonèmes étaient identifiés sur un mode catégoriel (présent/absent), comme dans les modèles de perception de la parole les plus classiques, dans la mesure où les informations se rapportant à la structure détaillée du signal de parole ne pourraient pas dans un tel cas franchir la couche des détecteurs de phonème.

Rappelons en outre que TRACE est un modèle *parallèle* de la reconnaissance des mots, les informations fournies par le signal se diffusant en cascade d'un niveau de traitement à l'autre. Cela signifie que les détecteurs de mot sont mis en activation dès que le signal aboutit au réseau, bien avant, donc, que soient identifiés tous les phonèmes dont le mot-cible est composé. Le niveau d'activation de chaque détecteur de mot est alors continuellement remis à jour au fur et à mesure que le signal est traité par le réseau. Il en résulte que tous les indices phonétiques relatifs à l'identité du mot-cible peuvent exercer une influence sur le niveau lexical sitôt après avoir atteint la couche d'entrée.

3.2.2. Le modèle Cohort

Contrairement à TRACE, Cohort est un modèle verbal que Marslen-Wilson n'a pas cherché à formaliser au travers d'un programme d'ordinateur (voir cependant Gaskell, Hare et Marslen-Wilson, 1995). La reconnaissance d'un mot s'opère dans ce modèle en deux étapes. Dans un pre-

mier temps, sont activées les entrées lexicales dont la partie initiale coïncide avec celle du mot à reconnaître. Ces entrées lexicales viennent former la cohorte initiale. Dans un second temps, et tandis que s'accomplit le traitement du signal de parole, les mots candidats sont éliminés les uns après les autres de la cohorte, dès qu'ils cessent de correspondre avec le signal. Le mot-cible est considéré comme étant reconnu à partir du moment où il est le seul à figurer encore dans la cohorte.

Cohort partage avec TRACE trois propriétés fondamentales pour ce qui nous concerne ici. En premier lieu, les informations traitées par le modèle sont encodées sous une forme quantitative, chaque mot, par exemple, se caractérisant par un certain niveau d'activation, proportionnel à la fois à sa fréquence d'utilisation et à son degré de correspondance avec le signal d'entrée. Cette propriété donne à supposer que de petites variations dans la forme sonore du mot-cible peuvent avoir une influence dans l'exploration du lexique, comme cela est le cas pour TRACE. En second lieu, le modèle Cohort repose lui aussi sur l'hypothèse que l'exploration du lexique s'opère parallèlement à l'analyse du signal de parole. Ensuite, le niveau d'activation des mots contenus dans la cohorte est soumis à des variations continuelles au fur et à mesure que l'on avance dans le signal. C'est ainsi qu'il est possible d'expliquer le fait que le mot « send » soit identifié par l'auditeur aussitôt que celui-ci perçoit des premières traces de nasalité au sein de la voyelle. En troisième lieu, il est supposé, dans Cohort comme dans TRACE, que le signal de parole est d'abord converti par l'auditeur sous la forme d'un ensemble de traits.

À la différence de TRACE, cependant, Cohort conduit à rejeter l'idée selon laquelle la reconnaissance des mots parlés s'accomplit par l'intermédiaire d'une représentation infra-lexicale de type phonémique. Dans le modèle Cohort, les mots sont identifiés par une mise en correspondance directe entre traits et lexique. La mise en place de la cohorte initiale et la sélection du mot-cible au sein de cet ensemble s'opèrent à partir de la représentation en traits elle-même. Cette hypothèse n'a pas toujours prévalu dans Cohort, qui se rangeait à l'origine dans la famille des modèles phonémiques (Marslen-Wilson et Welsh, 1978). Il est intéressant de noter que cette modification a été introduite à la suite d'une série d'expériences sur le rôle de la coarticulation dans la reconnaissance des mots (voir par ex. Warren et Marslen-Wilson, 1987). Selon Marslen-Wilson, il est difficile de comprendre comment des micro-variations phonétiques telles que celles présentées par la voyelle dans « send » et « said » pourraient trouver un chemin jusqu'au lexique mental, s'il faut supposer qu'une séquence d'entités phonémiques discrètes vient s'interposer entre le signal et le lexique. Dans sa version présente, Cohort postule que les traits sont projetés sur le lexique de manière directe (sans niveau intermédiaire de représentation) et continue (au fur et à mesure que ces traits sont extraits du signal de parole).

Cohort possède un certain nombre de points communs avec le modèle LAFF *(Lexical Access from Features)* conçu par Ken Stevens et présenté par celui-ci dans différentes conférences (voir par ex. Stevens, 1986). Dans ce

modèle, l'identification lexicale s'accomplit à partir d'une matrice de *traits asynchrones*. Au lieu d'être assemblés en faisceaux correspondant chacun à un phonème, tel qu'on le suppose dans les modèles phonémiques, les traits sont considérés dans LAFF comme évoluant dans le temps de manière (semi)-indépendante. Cette hypothèse constitue un tournant théorique important, dans la mesure où la notion de trait dans sa définition classique est intimement liée à celle de phonème. Le tableau I illustre les différences entre approche phonémique et approche non phonémique de la reconnaissance des mots, en montrant sous quelle forme il est possible de se représenter le mot « pawn » selon chacune de ces deux approches.

TABLEAU I. — *Représentation lexicale du mot « pawn » dans une perspective segmentale conventionnelle (à gauche) et dans le modèle LAFF de Stevens (à droite). Adapté de Klatt (1989)*

Lexical representation for « pawn », in a conventional segmental approach (left), and in Stevens's LAFF model (right). Adapted from Klatt (1989)

	p	ɔ	n		p	ɔ	n
high	–	–	–	high		–	
low	–	+	–	low		+	
back	–	+	–	back		+	
nasal	–	+	–	nasal			+
spread glottis	+	–	–	spread glottis	+		
sonorant	–	+	+	sonorant	–		
voiced	–	+	+	voiced	–		
strident	–	–	+	strident			
consonantal	+	–	+	consonantal	+		+
coronal	–	–	+	coronal	–		+
anterior	+	–	+	anterior	+		+
continuant	–	+	–	continuant	–		–

À gauche figure une représentation conventionnelle, dans laquelle une valeur (binaire) est attribuée à chaque trait pour chaque phonème. Cette représentation spécifie ainsi que /p/ et /ɔ/ sont – nasal et /n/ + nasal. La nasalisation de la voyelle /ɔ/ (phénomène régulier en anglais comme nous l'avons vu) sera alors interprétée comme un effet de surface. On postulera que la voyelle est modifiée dans sa forme phonétique sous l'influence de la consonne nasale suivante. Selon cette hypothèse, l'auditeur se doit de remonter de la forme de surface de la voyelle à sa forme sous-jacente, antérieure à la mise en jeu de cet effet de contexte, avant de pouvoir identifier correctement le mot prononcé.

À droite apparaît la forme phonologique attribuée au mot « pawn » dans LAFF. Contrairement à l'approche conventionnelle, les traits sont ici définis en certains points du signal de parole seulement (voir par ex. *high*, *low*, et *back*, associés au noyau vocalique, par opposition à *consonantal*, *coronal*, *anterior*, etc., associés à l'attaque et à la coda). En outre, les valeurs se distribuent de manière non-linéaire à l'intérieur de la matrice. Le caractère partiellement nasalisé de /ɔ/ est ainsi symbolisé par la présence du trait + nasal à mi-chemin entre /ɔ/ et /n/. De la même manière, le trait – voiced prend place dans la transition entre /p/ et /ɔ/ (au lieu d'être aligné avec /p/), sachant que c'est dans cette transition que résident les principaux indices acoustiques permettant à l'auditeur de percevoir /p/ comme étant non voisé (*voice onset time,* transition de F1, etc.).

En postulant que les traits sont directement mis en relation avec le lexique dans la reconnaissance des mots, Cohort et LAFF conduisent à l'évidence à fournir des phénomènes de coarticulation une interprétation nouvelle, déjà esquissée dans TRACE mais qui reçoit ici une forme plus explicite. En bref, le terme de coarticulation ne désigne pas dans Cohort ou LAFF un *processus* (phonème modifié sous l'influence d'un autre phonème) mais une *structure temporelle*, c'est-à-dire le fait que les traits se distribuent non linéairement à l'intérieur d'un mot. Dans le mot « send », pour reprendre notre exemple de base, on ne considérera pas que la voyelle /e/ est nasalisée sous l'influence de la consonne /n/, mais plutôt que le trait nasal prend la valeur + à l'intérieur d'un intervalle commun à la voyelle et à la consonne. Sur le plan de la perception, cette solution présente des avantages certains pour l'auditeur. Dans un modèle phonémique tel que TRACE, la reconnaissance des mots fait nécessairement suite à un processus complexe consistant à reconstruire la forme sous-jacente (invariante) de chaque phonème à partir de sa forme de surface (assujettie, elle, à l'influence du contexte). Dans un modèle non phonémique, la mise en relation entre traits et lexique possède un caractère beaucoup plus simple, chaque unité lexicale étant de surcroît supposée revêtir la forme d'une matrice de traits asynchrone, à l'image de la représentation de base.

3 . 3. DISCUSSION

Aussi fondamentales soient-elles du point de vue théorique, il est pour le moins difficile de faire ressortir les différences entre modèles phonémiques et modèles non phonémiques sur le plan empirique. La plupart des données expérimentales dont nous avons fait état ici se montrent compatibles avec l'une et l'autre de ces deux catégories de modèle. Selon Marslen-Wilson et Warren (1994), le fait que la sélection des candidats lexicaux soit dans la reconnaissance d'un mot sensible de manière continue aux variations temporelles présentées par le signal de parole (comme cela a été établi dans les expériences de type *gating*), va davantage dans le sens d'un modèle non phonémique tel que Cohort dans sa version présente. Comme nous

l'avons indiqué cependant, de tels résultats restent explicables dans le cadre d'un modèle phonémique de type TRACE, pour autant : *a)* que ce modèle soit un modèle parallèle du traitement de la parole ; et *b)* que l'information relative au mot-cible y soit représentée sous la forme d'un ensemble de paramètres continus.

On peut cependant citer un petit nombre d'études dont les résultats apportent aux modèles non phonémiques un soutien plus univoque. L'expérience de Streeter et Nigro (1979, voir section 2.2), en premier lieu, a fait apparaître qu'un auditeur réagit de manière différente à la présence d'un *mismatch* phonétique au sein d'un stimulus dissyllabique selon le statut lexical de ce stimulus. Lorsque les indices fournis par la première et la seconde voyelle sur la consonne médiane étaient rendus divergents par *cross-splicing*, les temps de réponse obtenus dans une tâche de décision lexicale présentaient un allongement pour les mots mais pas pour les non-mots. Selon Streeter et Nigro un tel patron de réponse donne à penser que les indices acoustiques situés de part et d'autre du point de *cross-splicing* font l'objet d'une intégration perceptive au niveau lexical seulement. Si l'on supposait en effet que le *mismatch* phonétique était détecté à un niveau infra-lexical, ce *mismatch* devrait entraîner un allongement du TR, que le stimulus soit un mot ou un non-mot. Streeter et Nigro en déduisent que l'identification d'un mot s'opère directement à partir d'une représentation phonétique globale de ce mot. Notons cependant que l'absence de différence dans les TR entre stimuli de contrôle et stimuli transcollés pour les non-mots pourrait également s'interpréter comme un simple effet de plafond, les non-mots ayant donné lieu en moyenne à des TR sensiblement plus longs que les mots.

Marslen-Wilson et Warren (1994, voir 2.2) ont récemment repris le paradigme expérimental employé par Streeter et Nigro en lui apportant d'importantes améliorations. Les stimuli utilisés dans cette expérience formaient des paires de type mot/non-mot telles que « job/smob ». Chaque mot existait en trois versions différentes, construites en adjoignant par transcollage à la consonne finale de ce mot une séquence CV tirée soit d'une autre répétition de ce même mot, soit d'un autre mot (ex. « jog »), soit enfin d'un non-mot (ex. « jod »). Selon le même principe, chaque non-mot se présentait lui aussi en trois versions. Le design de l'expérience est détaillé dans le tableau II, construit sur le modèle de la table 1 de Marslen-Wilson et Warren (1994).

Selon Marslen-Wilson et Warren, c'est en ce qui concerne les *non-mots* que les modèles phonémiques et non phonémiques conduisent à des prédictions divergentes. Dans un modèle phonémique, l'intégration des indices acoustiques situés avant et après le point de rupture va s'opérer à un niveau infra-lexical, alors que le signal d'entrée est converti en une chaîne de phonèmes. Par suite, on s'attendra à ce qu'un *mismatch* phonétique se traduise par un TR plus long dans une tâche de décision lexicale (par comparaison avec la condition de contrôle N1N1, selon la notation adoptée dans le tableau II), quelle que soit la manière dont le *mismatch* aura été engendré (M2N1

TABLEAU II. — *Design expérimental utilisé par Marslen-Wilson et Warren (1994)*

Experimental conditions in Marslen-Wilson and Warren (1994)

Type de séquence	Notation	Exemple
Mots		
1. mot 1 + mot 1	M1M1	*job* + j*ob*
2. mot 2 + mot 1	M2M1	*jog* + j*ob*
3. non-mot 3 + mot 1	N3M1	*jod* + j*ob*
Non-mots		
1. non-mot 1 + non-mot 1	N1N1	*smob* + sm*ob*
2. mot 2 + non-mot 1	M2N1	*smog* + sm*ob*
3. non-mot 3 + non-mot 1	N3N1	*smod* + sm*ob*

ou N3N1). Dans un modèle non phonémique tel que Cohort, en revanche, les non-mots de type M2N1 et de type N3N1 doivent donner à observer des patrons de réponse différents. Lorsqu'un non-mot de type M2N1 est présenté au sujet, on peut supposer que la partie initiale de ce non-mot (« smo(g) ») va être mise en correspondance avec le lexique pour activer un certain nombre d'unités lexicales (« smog » et ses compétiteurs). À la fin de la voyelle, la séquence sera interprétée comme un mot se terminant par une consonne vélaire. Lorsque la consonne finale (/b/) est entendue, une rupture de correspondance avec l'unité lexicale activée sera détectée, en donnant lieu alors à un délai dans la réponse du sujet. En revanche, lorsque le non-mot présenté au sujet est de type N3N1, la partie initiale de la séquence ne devrait pas être mise en relation avec une représentation lexicale. Par conséquent, le *mismatch* ne devrait pas entraîner de perturbation dans le traitement de la séquence, et la réponse produite par le sujet devrait être aussi rapide que dans la condition de contrôle. En d'autres termes, Cohort amène à prédire qu'un non-mot sera traité de manière différente par l'auditeur selon que la partie initiale de la séquence (avant le point de transcollage) provient elle-même d'un mot ou d'un non-mot.

Les résultats expérimentaux obtenus par Marslen-Wilson et Warren sont en accord avec leurs prédictions. Les auteurs montrent en outre que TRACE ne leur a pas permis de simuler ces résultats de manière satisfaisante. Pour être plus précis, TRACE répondait différemment aux stimuli de type mot selon la façon dont ils avaient été construits (performance moins bonne pour M2M1 que pour N3M1), alors que les sujets ne s'étaient pas montrés sensibles à de telles différences. Les données expérimentales et les simulations de Marslen-Wilson et Warren sont discutées en détail dans Norris, McQueen et Cutler (2000).

Les données de Hawkins et Nguyen (1999, voir 2.2) se prêtent également à une interprétation de type non phonémique. Rappelons qu'il a été établi dans cette expérience qu'un /l/ en position d'attaque dans un mot monosyllabique est produit différemment selon le caractère voisé ou non voisé de l'occlusive finale : /l/ est à la fois plus long et plus sombre quand l'occlusive est voisée plutôt que non voisée. L'expérience a fait apparaître que l'auditeur est dans certains cas sensible à ces petites variations dans la structure acoustique de /l/, lorsqu'il s'agit d'identifier le mot-cible. Selon les auteurs, ces variations contribuent à renforcer deux propriétés perceptives majeures associées au trait de voisement : le rapport de durée C :V *(C :V duration ratio)*, et la propriété basse-fréquence *(Low-Frequency property)*, pour reprendre les termes proposés par Kingston et Diehl (1994). Le rapport C :V désigne la durée de la tenue de l'occlusion rapportée à celle de la voyelle précédente, les occlusives voisées se caractérisant par un rapport plus petit que les non voisées. La propriété BF est à mettre en relation avec la présence d'énergie en basse fréquence au voisinage de la frontière entre voyelle et consonne pour les occlusives voisées (fréquence de F_1 plus basse en fin de voyelle, voisement pendant la tenue de l'occlusion). Selon Hawkins et Nguyen (1999), un /l/ d'attaque plus long devant une coda voisée pourrait contribuer à réduire davantage encore le rapport C :V, sous réserve que ce rapport soit légèrement redéfini de façon à prendre en compte la durée de l'attaque syllabique. Par ailleurs, le timbre relativement sombre du /l/ d'attaque devant une coda voisée pourrait venir renforcer la propriété basse-fréquence, dans la mesure où cet assombrissement se manifeste comme nous l'avons indiqué par un F_2 moins élevé. Ainsi, les caractéristiques acoustiques du /l/ d'attaque semblent se combiner avec celles des segments phonétiques suivants pour mettre en relief différentes propriétés perceptives reliées au voisement de la coda. L'important pour ce qui nous concerne ici est que le trait de voisement est à mettre en relation avec un ensemble d'indices distribués *sur toute la durée* de la syllabe (liquide initiale, voyelle et occlusive finale).

Nguyen et Hawkins (1999) soulignent que la version actuelle de TRACE ne permet pas d'expliquer la sensibilité de l'auditeur à ces indices acoustiques distribués sur de longs intervalles de temps. TRACE présuppose en effet que les indices rattachés à un contraste phonologique donné (ex. voisé / non voisé) sont contenus à l'intérieur d'un intervalle temporel de courte durée. Comme cela est rappelé plus haut, chaque détecteur de trait opère sur une fenêtre de 5 ms, et chaque détecteur de phonème sur une fenêtre de 15 ms. Par ailleurs, le contexte pris en compte dans l'identification d'un phonème-cible est confiné au phonème précédent et au phonème suivant (voir *supra*). Sans doute serait-il possible de simuler le rôle du /l/ d'attaque dans la perception du voisement de la coda en établissant des connexions entre consonnes d'attaque et coda. Cependant, en plus de faire considérablement augmenter le nombre de degrés de liberté du modèle (au détriment de son pouvoir explicatif), ces connexions conduiraient probablement à rendre TRACE un peu moins segmental et phonémique, dans la

mesure où la structure des syllabes (attaque, noyau, coda, etc.) aurait à y être représentée de manière explicite. Plus généralement, le fait qu'un trait distinctif puisse être rattaché à des indices acoustiques répartis en différents points dans la syllabe nous semble difficilement compatible avec un modèle phonémique de type TRACE. On voit mal, en effet, pourquoi des indices distribués dans le temps devraient être mis en relation avec des détecteurs à fenêtre temporelle courte (les détecteurs de phonème), si cette information doit être redistribuée à nouveau à partir de ces détecteurs vers des unités de niveau supérieur plus longues (les entrées lexicales). L'existence de liens à distance entre indices relatifs à un même trait distinctif entre ainsi en contradiction avec les modèles de type TRACE, en laissant supposer que le découpage du signal de parole en unités phonémiques n'est pas essentiel à la reconnaissance du mot-cible. Nguyen et Hawkins (1999) présentent une série d'arguments en faveur d'un modèle non phonémique basé sur l'hypothèse selon laquelle : 1 / le traitement de la parole fait intervenir deux fenêtres temporelles (une fenêtre courte pour les événements acoustiques rapides et une fenêtre longue pour les propriétés acoustiques distribuées) ; et 2 / la reconnaissance d'un mot s'opère en projetant directement une représentation phonétique détaillée du signal d'entrée sur le lexique.

4. CONCLUSION

On a longtemps considéré que la mise en correspondance entre signal de parole et lexique s'opérait à partir d'une représentation phonologique abstraite, prenant par exemple la forme d'une matrice de traits binaires. Cette hypothèse prévaut encore dans de nombreux modèles de la reconnaissance des mots. Les expériences que nous avons rapportées font pourtant apparaître que l'auditeur peut se montrer sensible dans l'identification des mots à de petites variations quantitatives dans la structure du signal de parole. Ces travaux donnent à penser que les représentations d'entrée dans l'accès au lexique sont des représentations très détaillées et que l'auditeur peut tirer parti des contrastes phonétiques les plus ténus pour différencier un mot de ses compétiteurs.

Au fil de cette revue des travaux, notre point de vue s'est en outre progressivement élargi. De cette question initiale relative à la structure phonétique des représentations d'entrée, l'accent s'est déplacé ensuite sur des problèmes de nature plus générale touchant à la mise en relation de la représentation d'entrée avec le lexique. À travers une comparaison établie entre TRACE et Cohort, nous avons montré ainsi que les effets de coarticulation, loin de constituer un phénomène de surface d'une importance secondaire dans la communication parlée, sont en fait susceptibles de nous fournir des indications essentielles sur les mécanismes cognitifs mis en œuvre dans l'identification lexicale. Selon de récentes expériences, la manière dont ces effets de coarticulation sont traités par l'auditeur laisse en particulier

supposer que l'identification d'un mot s'accomplit directement à partir de la représentation d'entrée, sans passer par l'intermédiaire d'une représentation infra-lexicale de type phonémique. Les effets de coarticulation occupent ainsi une place centrale dans le débat entre modèles phonémiques et non phonémiques de la reconnaissance des mots.

Réciproquement, les études expérimentales réalisées aujourd'hui sur la compréhension du langage oral sont à l'évidence en mesure de jeter une lumière nouvelle sur la structure d'un système phonétique. Le rôle de la coarticulation dans la reconnaissance des mots nous a ainsi conduit à remettre en question l'hypothèse classique selon laquelle le signal de parole est analysé par l'intermédiaire d'un ensemble de détecteurs de phonèmes. Les expériences de type *gating* tendent également à faire prévaloir l'idée que les traits phonétiques se distribuent de manière non-linéaire à l'intérieur de la chaîne parlée, en empiétant les uns sur les autres dans le domaine temporel. Ces travaux menés dans le domaine de la reconnaissance des mots sont de nature à transformer la manière dont un système phonétique peut être caractérisé. Ils nous conduisent à aborder l'analyse des systèmes phonétiques dans une perspective nouvelle, en les rapportant à la fonction première qui est la leur, pour l'auditeur, de véhiculer du sens.

RÉSUMÉ

Dans cet article, nous passons en revue un ensemble d'expériences récemment réalisées sur le rôle des effets de coarticulation dans la reconnaissance des mots. En désaccord avec une hypothèse qui a longtemps prévalu, ces travaux donnent à penser que la reconnaissance d'un mot ne s'opère pas à partir d'une représentation phonologique abstraite du signal de parole. Les données les plus récentes font au contraire apparaître que l'auditeur peut, dans l'identification d'un mot, être directement influencé par une grande variété de contrastes phonétiques extrêmement ténus. Les implications de ces résultats en ce qui concerne les modèles actuels de la reconnaissance des mots sont discutées.

Mots-clés : coarticulation, reconnaissance des mots, perception de la parole.

BIBLIOGRAPHIE

Ali L., Gallagher T., Goldstein J., Daniloff R. G. — (1971) Perception of coarticulated nasality, *Journal of the Acoustical Society of America*, *49*, 538-540.

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

Browman C., Goldstein L. — (1986) Towards an articulatory phonology, *Phonology Yearbook*, *3*, 219-252.

Byrd D. — (1992) Perception of assimilation in consonants clusters : A gestural model, *Phonetica*, *49*, 1-24.

Daniloff R. G., Hammarberg R. E. — (1973) On defining coarticulation, *Journal of Phonetics*, *1*, 239-248.

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

Fowler C. A. — (1980) Coarticulation and theories of extrinsic timing, *Journal of Phonetics*, *8*, 113-133.

Fowler C. A. — (1984) Segmentation of coarticulated speech in perception, *Perception et Psychophysics*, *36*, 359-368.

Fowler C. A., Smith M. R. — (1986) Speech perception as « vector analysis » : An approach to the problems of invariance and segmentation, *in* J. S. Perkell et D. H. Klatt (Edit.), *Invariance and variability in speech processes*, Hillsdale (NJ), Lawrence Erlbaum, 123-136.

Frauenfelder U. H. — (1991) Une introduction à la reconnaissance des mots parlés, *in* R. Kolinsky, J. Morais et J. Segui (Édit.), *La reconnaissance des mots dans différentes modalités sensorielles. Données et modèles en psycholinguistique cognitive*, Paris, PUF, 7-36.

Frauenfelder U. H. — (1992) The interface between acoustic-phonetic and lexical processing, *in* M. E. H. Schouten (Edit.), *The auditory processing of speech : From sounds to words*, Berlin, Mouton de Gruyter.

Frauenfelder U., Nguyen N. — (2000) Le traitement du langage oral, *in* J. Rondal et J. A. Seron (Édit.), *Troubles du langage : bases théoriques, diagnostic et rééducation*, Bruxelles, Mardaga, 213-240.

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

Gaskell G., Marslen-Wilson W. — (1993) *Match and mismatch in phonological context*, Proceedings of the 15th Annual Conference of the Cognitive Science Society, Hillsdale (NJ), Lawrence Erlbaum.

Gaskell M. G., Hare M., Marslen-Wilson W. D. — (1995) A connectionist model of phonological representation in speech perception, *Cognitive Science*, *19*, 407-439.

Gaskell M. G., Marslen-Wilson W. — (1996). Phonological variation and inference in lexical access, *Journal of Experimental Psychology : Human Perception and Performance*, *22*, 144-158.

Grosjean F. — (1980) Spoken word recognition processes and the gating paradigm, *Perception and Psychophysics*, *36*, 267-283.

Gupta P., Mozer M. C. — (1993) *Exploring the nature and development of phonological representations*, Proceedings of the 15th Annual Conference of the Cognitive Science Society.

Harris K. S. — (1958) Cues for the discrimination of American English fricatives in spoken syllables, *Language and Speech*, *1*, 1-7.

Hawkins S., Nguyen N. — (1999) Effects on word recognition of syllable-onset cues to syllable-coda voicing, *in* J. Local (Edit.), *Papers in laboratory phonology VI*, Cambridge (UK), Cambridge University Press, à paraître.

Ingram J., Mylne T. — (1994) *Perceptual parsing of nasal vowels*, Proceedings of ICSLP 94, 495-498.

Katz W. F., Kripke C., Tallal P. — (1991) Anticipatory coarticulation in the speech of adults and young children : Acoustic, perceptual, and video data, *Journal of Speech and Hearing Research*, *34*, 1222-1232.

Kearns R., Nguyen N. — (1997) *The perception of vowel nasalization in English-French bilinguals*, International Symposium on Bilingualism, Newcastle upon Tyne, 9-12 avril 1997.

Kingston J., Diehl R. L. — (1994) Phonetic knowledge, *Language*, *70*, 419-454.

Klatt D. H. — (1988) Review of selected models of speech perception, *in* W. D. Marslen-Wilson (Edit.), *Lexical representation and process*, Cambridge (MA), MIT Press, 169-226.

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

Local J. — (1992) Modelling assimilation in a non-segmental, rule-free synthesis, *in* G. J. Docherty et D. R. Ladd (Edit.), *Papers in laboratory phonology II*, Cambridge (UK), Cambridge University Press, 190-223.

Mann V. A., Repp B. H. — (1980) Influence of vocalic context on perception of the [ʃ]-[s] distinction, *Perception and Psychophysics*, *28*, 213-228.

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

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

Marslen-Wilson W. D., Gaskell M. G. — (1992) Match and mismatch in lexical access, *International Journal of Psychology*, *27*, 61.

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

Martin J. G., Bunnell H. T. — (1981) Perception of anticipatory coarticulation effects, *Journal of the Acoustical Society of America*, *69*, 559-567.

Martin J. G., Bunnell H. T. — (1982) Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences, *Journal of Experimental Psychology : Human Perception et Performance*, *8*, 473-488.

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

Nguyen N. — (1995) *Contextual and lexical effects in the identification of fricatives*, Proceedings of the XIIIth Congress of Phonetic Sciences, Stockholm, vol. 2, 530-533.

Nguyen N. — (1999) *La coarticulation : aspects articulatoires, acoustiques et perceptifs*, Mémoire d'habilitation à diriger des recherches, Université Lumière - Lyon 2, Lyon.

Nguyen N., Hawkins S. — (1998) *Syllable-onset acoustic properties associated with syllable-coda voicing*, Proceedings of the 5th International Conference on Spoken Language Processing, Sydney, Australie.

Nguyen N., Hawkins S. — (1999) *Implications for word recognition of phonetic dependencies between syllable codas and onsets*, Symposium on the perception of coarticulated speech, XVIth International Congress of Phonetic Sciences, San Francisco.

Nguyen N., Wrench A., Gibbon F., Hardcastle W. J. — (1998) *Articulatory, acoustic and perceptual aspects of fricative/stop coarticulation*, Proceedings of the 5th International Conference on Spoken Language Processing, Sydney, Australie.

Nix A., Gaskell G., Marslen-Wilson W. — (1993) *Phonological variation and mismatch in lexical access*, Proceedings of Eurospeech '93, Berlin, vol. 1, 685-688.

Nolan F. J. — (1992) The descriptive role of segments : Evidence from assimilation, *in* G. J. Docherty et D. R. Ladd (Edit.), *Papers in laboratory phonology II : Gesture, segment, prosody*, Cambridge (UK), Cambridge University Press, 261-280.

Norris D. — (1992) Connectionism : A new breed of bottom-up model, *in* R. Reilly et N. Sharkey (Edit.), *Connectionist approaches to natural language processing*, Hove (UK), Lawrence Erlbaum.

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

Ohala J., Ohala M. — (1995) Speech perception and lexical representation : The role of vowel nasalization in Hindi and English, *in* B. Connell et A. Arvaniti (Edit.), *Papers in laboratory phonology, 4 : phonology and phonetic evidence*, Cambridge (UK), Cambridge University Press, 41-60.

Öhman S. — (1966) Coarticulation in VCV utterances : Spectrographic measurements, *Journal of the Acoustical Society of America*, *39*, 151-168.

Ostreicher H. J., Sharf D. J. — (1976) Effects of coarticulation on the identification of deleted consonants and vowel sounds, *Journal of Phonetics*, *4*, 285-301.

Perkell J. S., Klatt D. H. (Edit.) — (1986) *Invariance and variability in speech processes*, Hillsdale (NJ), Lawrence Erlbaum.

Pitt M. A., Samuel A. G. — (1993) An empirical and meta-analytic evaluation of the phoneme identification task, *Journal of Experimental Psychology : Human Perception and Performance*, *19*, 699-795.

Pols L. C. W., Schouten M. E. H. — (1978) Identification of deleted consonants, *Journal of the Acoustical Society of America*, *64*, 1333-1337.

Serniclaes W., Beeckmans R., Radeau M. — (1995) Phonetic and lexical effects in speech perception, *in* C. Sorin, J. Mariani, H. Méloni et J. Schoentgen (Edit.), *Levels in speech communication : Relations and interactions*, Amsterdam, Elsevier, 39-50.

Shillcock R., Levy J., Lindsey G., Cairns P., Chater N. — (1993) Connectionist modelling of phonological space, *in* T. M. Ellison et J. M. Scobbie (Edit.), *Computational phonology, Edinburgh working papers in cognitive science*, vol. 8, 179-195.

Shockey L., Watkins A. — (1995) *Reconstruction of base forms in perception of casual speech*, Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, vol. 3, 588-591.

Sotillo C., McAllister J., Bard E. G., Doherty-Sneddon G., Newlands A. — (1995) *Word intelligibility and place assimilation in spontaneous speech*, Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, vol. 2, 550-553.

Stevens K. N. — (1986) *Models of phonetic recognition II : A feature-based model of speech recognition*, Proceedings of the Montreal Satellite Symposium on Speech Recognition, XIIth International Congress on Acoustics, 66-67.

Streeter L. A., Nigro G. N. — (1979) The role of medial consonant transitions in word perception, *Journal of the Acoustical Society of America*, *65*, 1533-1541.

Van Heuven V. J., Jongenburger W. — (1993) *Perceptual effects of place and voicing assimilation in Dutch consonants*, Proceedings of Eurospeech '93, Berlin, vol. 2, 1503-1506.

Van Santen J. P. H., Coleman J. S., Randolph M. A. — (1992) *Effects of postvocalic voicing on the time course of vowels and diphthongs*, Journal of the Acoustical Society of America, *92*, 2444.

Warren P., Marslen-Wilson W. — (1987) Continuous uptake of acoustic cues in spoken word recognition, *Perception and Psychophysics*, *41*, 262-275.

Warren P., Marslen-Wilson W. — (1988) Cues to lexical choice – discriminating place and voice, *Perception and Psychophysics*, *43*, 21-30.

Whalen D. H. — (1984) Subcategorical phonetic mismatches slow phonetic judgments, *Perception and Psychophysics*, *35*, 49-64.

Whalen D. H. — (1991) Subcategorical phonetic mismatches and lexical access, *Perception and Psychophysics*, *50*, 351-360.

Winitz H., Scheib M. E., Reeds J. A. — (1971) Identification of stops and vowels from the burst portion of /p,t,k/ isolated from conversational speech, *Journal of the Acoustical Society of America*, *51*, 1309-1317.

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

(Accepté le 20 septembre 1999.)