Traitement phonétique et représentation lexicale dans la reconnaissance des mots

SONIA KANDEL & LOUIS-JEAN BOË

Institut de la Communication Parlée, URA - CNRS n° 368, INPG - Université Stendhal, Domaine Universitaire, BP 25 - 38040 Grenoble cedex 9, France

ABSTRACT

The aim of very many studies has been the clarification of the process of word recognition. The scope of the present study is limited to the description of the COHORT model, proposed by Marslen-Wilson and his colleagues since 1978, and of the SARAH model, postulated by the proponents of the syllabic hypothesis (Mehler, Segui, and their colleagues). These authors develop original approaches that present the advantage to integrate the problems related with the processing of the acoustic signal and the word representation, and this in several languages.

RÉSUMÉ

De très nombreuses études se sont données pour objectif l'éclaircissement du processus de reconnaissance des mots. Nous nous limiterons à la description du modèle COHORT, proposé par Marslen-Wilson et ses collaborateurs depuis 1978, et du modèle SARAH, postulé par les tenants de l'hypothèse syllabique (Mehler, Segui et leurs collaborateurs). Ces auteurs développent des approches originales qui ont l'avantage d'intégrer les problèmes liés au traitement du signal acoustique et à la représentation des mots, et cela en plusieurs langues.

1. Introduction

Les processus mis en œuvre pour comprendre le message d'un interlocuteur sont d'une extrême complexité. La compréhension de la parole implique non seulement le décodage d'un message linguistique, mais aussi la mise en relation de ce message codifié avec un vaste répertoire d'informations de natures diverses. Le haut degré d'automaticité des mécanismes impliqués ne permet pas à l'auditeur d'être conscient des processus intermédiaires mais seulement du résultat terminal du traitement linguistique. Ce processus complexe est souvent décomposé en une série de sous-processus distincts et fonctionnellement indépendants, l'attention porte sur un aspect particulier :

Bulletin de la communication parlée n° 3, p. 61-72.

les processus de reconnaissance des mots. Est avancée l'hypothèse selon laquelle la compréhension d'une chaîne parlée requiert, entre autres, la reconnaissance des mots qui la composent. La reconnaissance d'un mot consisterait en un appariement entre une représentation construite à partir de l'information sensorielle décodée et une représentation de ce mot stockée dans le cerveau de l'interlocuteur ; ce processus correspond aux toutes premières millisecondes de la perception de la parole.

Les travaux sur la reconnaissance des mots ont fait émerger le concept de *lexique mental* (ou *lexique interne* ou *lexique subjectif*) : la composante du système de traitement du langage qui a trait aux connaissances que le locuteur et son interlocuteur partagent pour les mots de la langue avec laquelle ils communiquent (Segui, 1991). Le processus de reconnaissance de mots commence par l'activation des mécanismes perceptifs par une entrée sensorielle. L'interlocuteur décode l'information reçue et en construit une représentation, la *représentation d'entrée.* Celle-ci est comparée avec les *représentations lexicales* – répertoriées dans le lexique mental – jusqu'à ce qu'un appariement entre la représentation d'entrée et la représentation lexicale soit satisfaisant. L'identité entre les deux représentations détermine la reconnaissance du mot. Frauenfelder (1991) définit deux opérations de base dans la reconnaissance des mots : l'*identification lexicale*, qui consiste à localiser la description de la forme ou l'adresse d'une entrée-cible dans le lexique mental, et l'*accès lexical* impliquant les différents types d'informations (sémantiques et syntaxiques) associées à cette entrée.

De très nombreuses études se sont données pour objectif la compréhension des mécanismes perceptifs, la nature des indices utilisés et le mode de traitement de l'information par le système de reconnaissance qui permet l'appariement satisfaisant dans le processus d'accès au lexique.

Les questions soulevées concernent essentiellement :
1. la nature du mécanisme de décodage de l'information sensorielle et le format sous lequel l'information est encodée ;
2. les processus qui rendent possible l'accès aux représentations lexicales ;
3. le format des représentations lexicales ;
4. l'organisation du lexique mental.

Des réponses de natures diverses ont été données à ces questions. Nous focaliserons notre attention sur les problèmes relatifs au traitement du signal acoustique et à la représentation lexicale impliqués dans le processus de reconnaissance des mots. L'objectif de cet article n'est pas de présenter une liste exhaustive des modèles ni de les comparer entre eux, mais plutôt de donner au lecteur une idée générale sur les problèmes de modélisation rencontrés dans l'étude sur la reconnaissance des mots parlés. Nous proposerons donc une description de deux modèles – COHORT et SARAH – qui intègrent des notions sur le traitement phonétique et sur la représentation lexicale. Le choix de ces deux modèles repose principalement sur le fait qu'ils ont été élaborés à partir d'un vaste ensemble d'expériences réalisées en plusieurs langues et que l'on peut observer leur évolution en fonction des critiques théoriques et/ou méthodologiques soulevées par des résultats expérimentaux destinés à les tester.

La problématique inhérente à la description du processus de reconnaissance des mots est directement liée aux caractéristiques intrinsèques de la parole. La parole est variable : les gestes articulatoires n'échappent pas à la nature des mouvements humains ; leur faible reproductibilité est à l'origine de l'impossibilité de produire deux chaînes parlées identiques. La coarticulation, le

débit d'élocution et les différences intra et interlocuteur ont pour conséquence qu'un même mot est toujours réalisé sous des formes différentes. La reconnaissance du mot implique donc la convergence de nombreux exemplaires vers une représentation lexicale unique. Le caractère essentiellement continu de la parole pose également des problèmes de description. À la différence de l'écriture, la chaîne parlée ne présente pas de séparateurs pour indiquer la fin d'un mot et/ou le début du suivant. Le discours doit donc être segmenté en unités discrètes. Ce processus de *segmentation* de la parole – la constitution de l'unité-minimale de décodage sensoriel à partir d'un signal acoustique – est le noyau du processus de traitement lexical. Il implique l'extraction préalable d'unités sub-lexicales (les traits phonétiques, les phonèmes, les syllabes) dans un format qui soit lisible par le système de reconnaissance. La taille de l'unité de base est fondamentale pour déterminer le moment de déclenchement du mécanisme de recherche lexicale et pour la description des caractéristiques de la représentation lexicale.

Sur ces critères, Mehler *et al.* (1990) distinguent deux types de modèles :
1. les modèles *fine-grained* proposent un processus de traitement qui postule la continuité du flux de l'information. Le système traite l'information acoustique en temps réel, ce qui lui permet de sélectionner le meilleur candidat dès qu'il dispose de suffisamment d'information. Le modèle COHORT révisé (Marslen-Wilson, 1987) propose un système de traitement lexical faisant intervenir des unités minimales relativement fines – des traits distinctifs – qui sont directement transmises au niveau des représentations lexicales, sans le passage préalable par un niveau de traitement intermédiaire ;
2. les modèles *coarse-grained* présupposent un flux d'information discontinu entre les différents niveaux du traitement lexical, celui-ci étant dépendant d'unités de traitement relativement larges. Un niveau intermédiaire de traitement, le niveau pré-lexical, accumule l'information acoustique émise avant de la faire transiter vers les niveaux supérieurs. Une quantité critique d'informations périphériques doit être traitée pour le déclenchement du processus de reconnaissance. Ainsi, Mehler *et al.*, (1990) proposent le modèle SARAH, avec une unité pré-lexicale semblable à la syllabe.

Les deux modèles que nous présentons dans cet article illustrent la problématique inhérente aux caractéristiques de l'unité minimale de traitement ainsi que différentes approches concernant le format des représentations lexicales. Le premier nous renseigne davantage sur les problèmes liés aux processus de traitement de l'information acoustique alors que le deuxième est plus centré sur la description et acquisition des représentations. Ces différences d'approche les rendent, en quelque sorte très complémentaires.

2. Deux modèles de reconnaissance des mot

2.1 Le modèle COHORT

2.1.1 Marslen-Wilson & Welsh (1978)

COHORT a été le premier modèle de reconnaissance des mots appliqué aux mots parlés et à être configuré par rapport aux propriétés du signal acoustique. En effet, les tous premiers modèles pour la reconnaissance des mots ont concerné les mots écrits. La première version du modèle COHORT a été présentée par Marslen-Wilson & Welsh en 1978. Cette version initiale suppose deux étapes

successives pour la reconnaissance d'un mot. D'abord, la phase de constitution de la cohorte initiale, dans laquelle toutes les représentations lexicales s'appariant à la représentation d'entrée – représentation de nature phonémique – sont activées simultanément. L'alignement, qui met en correspondance des éléments de la représentation d'entrée à ceux de la représentation lexicale, se fait strictement à partir du début de la chaîne ; seule la partie initiale du mot pouvant générer des membres de la cohorte. Dans la deuxième étape, l'état de chaque représentation lexicale se modifie en fonction de la qualité de l'ajustement avec la représentation d'entrée. Le nombre de représentations lexicales dans la cohorte est progressivement réduit, lorsque la quantité d'information sur le mot-cible augmente la décision – par tout ou rien – sur l'identité du mot-cible est déterminante : elle est prise sur la base d'un appariement rigoureux entre la représentation d'entrée et la représentation lexicale. La reconnaissance du mot a lieu lorsqu'un seul mot reste dans la cohorte.

L'avantage de cette première version du modèle est qu'il permet de prédire le moment auquel le mot sera reconnu. Le point de reconnaissance correspond au *point d'unicité* : le moment à partir duquel la suite de sons devient unique par rapport aux autres chaînes initiales des mots du lexique. Empiriquement, il existe deux possibilités pour rendre le point d'unicité opérationnel. Malheureusement, toutes les deux reposent sur un certain nombre de décisions difficiles et arbitraires. La première est définie à l'aide d'un dictionnaire phonétique : le point d'unicité correspond au point à partir duquel aucune autre séquence initiale de phonèmes n'est partagée par d'autres mots dans le dictionnaire. La taille et le contenu du dictionnaire sont alors un facteur essentiel ; par exemple, l'inclusion ou l'exclusion des mots de même racine morphologique relativisent la localisation du point d'unicité. En outre, la coarticulation des gestes de production complique la précision de sa localisation. Le point de reconnaissance peut également être spécifié expérimentalement avec la technique du *fenêtrage* (*gating* : Grosjean, 1980 ; cf. *infra*). Cette procédure fournit une mesure temporelle sur le point de reconnaissance du mot, mais elle implique également des décisions arbitraires : l'inclusion des jugements de certitude des réponses, le pourcentage minimal de reconnaissance correcte et la durée des fenêtres.

L'importance du concept de point d'unicité repose sur son efficacité de prédiction observée dans les expériences sur la latence de reconnaissance des mots (cf. Frauenfelder *et al.*, 1990 ; Radeau & Morais, 1990). Cette notion suppose la discrimination de la représentation lexicale du mot-cible par rapport aux autres membres de la cohorte, sur la base de contraintes sensorielles et contextuelles. Marcus & Frauenfelder (1985) ont suggéré toutefois que l'effet du point d'unicité pourrait résulter de l'accroissement du nombre de phonèmes différents entre le mot-cible et les représentations lexicales après le point d'unicité, plutôt que de la réduction de la cohorte à un candidat lexical unique.

2.1.2 Marslen-Wilson (1987)

Il s'agit d'une version révisée qui tient compte d'un certain nombre de problèmes soulevés au niveau théorique aussi bien qu'expérimental. L'appariement par tout ou rien entre la représentation d'entrée et la représentation lexicale rend difficile, par exemple, l'explication de la reconnaissance des mots mal prononcés. Marslen-Wilson propose un mécanisme de décision basé sur la *qualité relative de l'ajustement.* Le modèle introduit le concept de *niveaux d'activation* où les mots ont des statuts différents en fonction de leur ajustement avec l'entrée : le niveau d'activation est d'autant plus élevé que l'ajustement entre la représentation d'entrée et la représentation lexicale est important. Le traitement repose sur un processus de *compétition* entre les représentations

lexicales activées. L'activation simultanée de multiples candidats lexicaux, l'émergence au cours du temps du meilleur candidat et la discrimination, deviennent possibles au fur et à mesure que le niveau d'activation pour le meilleur candidat atteint un niveau critique de différenciation par rapport aux niveaux d'activation de ses concurrents. Le déroulement temporel et le résultat du processus de reconnaissance reflètent non seulement l'activation candidat du mot-cible, mais aussi l'inactivation d'autres candidats, en particulier ses plus forts concurrents. La notion de niveau d'activation offre également la possibilité d'expliquer les effets de fréquence. C'est un phénomène dans lequel les mots d'usage plus fréquent dans la langue sont plus aisément restitués que les mots dont l'usage est plus rare : le mot *chapeau* par exemple, est généralement restitué plus vite que le mot *chameau*. La fréquence linguistique du mot détermine son taux d'activation : les mots les plus fréquents de la langue seront plus activés que les mots moins fréquents, le niveau d'activation des premiers étant susceptible d'augmenter plus rapidement que celui des derniers.

Dans cette nouvelle version, le traitement lexical ne repose plus sur une représentation phonémique, mais plutôt sur une représentation codifiée en termes de traits phonologiques. Ce changement s'appuie sur des études publiées par Warren & Marslen-Wilson (1987, 1988) dans lesquelles les auteurs mettent en évidence un processus de traitement faisant intervenir des traits distinctifs. Avant la fin d'une voyelle les sujets sont capables de distinguer entre le caractère plosif ou nasal de la consonne qui la suit (*flown-float*) et ils sont également sensibles de façon précoce au voisement (*mob-mop*) et au lieu d'articulation (*pat-pack*). Dans ces résultats on peut aussi noter que les courbes d'identification évoluent de manière continue en fonction de la quantité de signal présenté. Sur cette base, les auteurs concluent que le traitement de la parole est *continu* et ne repose sur aucun segment qui introduirait un délai ou discontinuité dans la reconnaissance des mots (cf. *supra* : modèle *fine-grained*).

Bien que cette version de 1987 soit plus réaliste pour le processus de traitement de l'information, elle affaiblit la capacité prédictive du modèle, notamment en ce qui concerne l'évolution temporelle de l'accès au lexique. En effet, la prédiction du point de reconnaissance se complique avec la considération de la qualité relative de l'ajustement, du niveau d'activation des candidats lexicaux et des caractéristiques structurelles des représentations lexicales.

2.1.3 Lahiri & Marslen-Wilson (1991, 1992)

En 1991, Lahiri & Marslen-Wilson ont proposé un modèle psycholinguistique qui intègre explicitement les processus de traitement et la nature des représentations lexicales impliqués dans la reconnaissance des mots. Plus spécifiquement, sont supposés des processus d'accès et de sélection déterminés par la structure des représentations lexicales. L'auditeur traite l'information acoustique sur la base d'une représentation lexicale abstraite et sous-spécifiée. Celle-ci contient des informations sur les segments et leurs traits phonologiques (« *integrated segmental-featural representation* ») qui correspondent à ceux des phonologies de Archangeli (1984) et Clements (1985). Les informations predictibles et non-distinctives sur ces traits ne sont pas incluses dans la représentation, d'où la notion de sous-spécification. Cette hypothèse a été testée à l'aide d'une étude comparative dans laquelle les auteurs ont pu observer qu'un même trait de nasalité est interprété en fonction des représentations lexicales sous-spécifiées, celles-ci étant déterminées par les descriptions phonologiques de chaque langue.

L'hypothèse de la sous-spécification dans l'accès au lexique postulée par Lahiri & Marslen-Wilson (1991, 1992) est mise en question par Ohala & Ohala (à paraître). Ces derniers affirment, également à partir de résultats obtenus dans une étude comparative, que les processus perceptifs

reposent sur l'information phonétique de surface plutôt que sur une représentation phonologiquement sous-spécifiée : l'interlocuteur base essentiellement ses décisions sur l'information qui lui est disponible dans le signal de parole, plutôt que sur des représentations auxquelles les indices acoustiques sont associés.

Les présupposés de Lahiri & Marslen-Wilson ainsi que ceux de l'étude de Ohala & Ohala ont été testés avec la technique du fenêtrage. L'utilisation de cette technique pour l'étude des représentations a été fortement critiquée, en particulier par McQueen (à paraître). Une brève description de cette tâche permettra au lecteur de mieux comprendre pourquoi. Une chaîne parlée est présentée avec des augmentations graduelles de 20 à 30 ms et le sujet doit désigner le mot qui, croit-il, est en train d'être prononcé. La réponse sur l'identité du mot-cible est accompagnée d'une échelle de certitude où il est demandé au sujet d'indiquer le degré de certitude de sa réponse. Selon les utilisateurs de cette technique, les mesures relevées sont des indicateurs du déroulement du processus de reconnaissance des mots : le temps de réponse permet d'observer l'établissement de l'information au cours du temps ; l'évaluation de la certitude de la réponse fournit des informations approximatives sur le niveau d'activation des représentations lexicales ; les erreurs permettent d'explorer la constitution de la cohorte initiale et d'identifier les représentations lexicales qui entrent en compétition dans le processus de reconnaissance.

McQueen (à paraître) insiste sur l'existence d'un biais de traitement dans ce type de tâche expérimentale : dans les premières fenêtres les sujets montrent une tendance à entendre un percept unifié (des mots monosyllabiques) plutôt que d'associer l'information perçue à une information absente (la fin du mot). D'après l'auteur, lorsque l'information acoustique n'est pas suffisante pour spécifier un mot, elle serait associée au candidat lexical le plus vraisemblable sur la base d'informations de nature acoustico-phonétique. La technique du fenêtrage serait donc plus adaptée à l'étude du processus de traitement phonétique qu'à celle des représentations lexicales car, selon McQueen (à paraître), elle ne serait pas en mesure de fournir des informations sur la nature des représentations ni sur leur influence sur l'analyse acoustico-phonétique.

2.2 Le modèle SARAH

2.2.1 L'hypothèse syllabique

Les tenants de l'hypothèse syllabique ont mené des très nombreuses recherches pour montrer l'importance de la syllabe dans la production et dans la perception de la parole, avant de proposer, en 1990, un modèle du processus de reconnaissance des mots : SARAH (« *Syllable Acquisition, Representation and Access Hypothesis* » ; Mehler, Dupoux & Segui). Nous porterons notre attention sur les aspects qui sont directement en rapport avec la modélisation du processus de reconnaissance des mots. L'hypothèse syllabique postule que des représentations sub-lexicales de type syllabique jouent un rôle essentiel dans la reconnaissance des mots. Après l'analyse acoustico-phonétique, la syllabe intervient à un niveau de traitement intermédiaire responsable du déclenchement du processus d'accès au lexique mental lorsque l'accumulation d'information sensorielle est suffisante. La syllabe ne doit pas être considérée comme une unité d'analyse du signal mais plutôt comme une unité d'organisation des différentes sources d'information acoustique possédant, au niveau de la forme (*Gestalt*) des propriétés d'invariance (Dupoux, 1989).

La plupart des études ayant trait à l'hypothèse syllabique ont été menées au moyen de la tâche de détection de phonèmes (« *phoneme monitoring* »), paradigme expérimental très différent à celui du fenêtrage. Dans celui-ci, une liste d'items est présentée au sujet qui doit répondre le plus

rapidement possible indiquant la présence d'un phonème-cible préalablement spécifié. La détection de la cible présuppose que le sujet ait accédé à son lexique mental pour donner sa réponse.

Dans l'approche syllabique, les phonèmes ne sont pas identifiés sur la base d'une unité plus fine (les traits distinctifs) mais ils sont dérivés d'une unité plus large, apparentée à la syllabe. Liberman *et al.*, (1974) ont montré que les enfants de moins de cinq ans ont des grandes difficultés à manipuler des phonèmes dans des tâches pour lesquelles il leur est demandé d'ajouter ou d'éliminer un phonème. Un phénomène similaire a été observé par Morais *et al.*, (1979) chez des adultes portugais illettrés, alors que ces mêmes sujets n'ont aucune difficulté à manipuler des syllabes (voir Morais *et al.*, 1991, pour une discussion sur ce sujet). La segmentation en phonèmes ne semble donc pas être accessible d'emblée, seul l'apprentissage d'un système d'écriture orthographique permettrait la prise de conscience des phonèmes dans la chaîne parlée. Ces résultats semblent indiquer que les phonèmes ne sont pas les segments de base de la perception, mais sont dérivés d'unités comme la syllabe. Ceci semble corroboré par Segui *et al.*, (1981) qui observent que les temps de détection des phonèmes et des syllabes sont fortement corrélés.

L'émergence d'unités intermédiaires de représentation entre le phonème et le mot – les syllabes et leurs constituants – peut permettre de limiter efficacement le nombre de candidats lexicaux à examiner lors du processus de reconnaissance. Mehler *et al.*, (1981) ont montré que l'identification des séquences phonétiques dans des mots dépend de leur structure syllabique : la séquence [pal] est détectée plus rapidement dans *pal-mier* que dans *pa-lace* et l'inverse est observé pour la syllabe [pa]. Or, la rapidité avec laquelle un sujet est capable d'identifier la partie initiale d'un mot dépend de l'organisation structurelle de la langue considérée. En effet, des sujets français segmentent la parole en unités de nature syllabique alors que des sujets anglais ne le font pas de la même façon (Cutler *et al.*, 1986). L'absence d'effet syllabique chez les sujets anglais a été expliquée par la nature ambisyllabique de certains segments : le [l] dans *balance* et *balcon* peut aussi bien appartenir à la première comme à la deuxième syllabe. Cette explication a été mise en question par Zwitserlood *et al.*, (1991) car en hollandais, langue dans laquelle l'on retrouve des segments ambisyllabiques, des effets syllabiques ont été observés. De plus, Rietveld & Frauenfelder (1987) ont montré qu'en hollandais l'effet syllabique se manifeste uniquement pour les consonnes liquides [l] et [r] mais n'apparaît pas avec les consonnes nasales et les occlusives orales.

Comme le souligne Frauenfelder (1992), les recherches dans d'autres langues que le français ont montré que plusieurs facteurs peuvent être à l'origine de l'absence d'effet syllabique : la place de la syllabe accentuée, la taille de la voyelle ainsi que la présence de réduction vocalique. Dupoux & Mehler (1992) suggèrent également que l'absence d'effet syllabique en japonais (Otake *et al.*, sous presse) et en espagnol (Sebastián-Gallés *et al.*, 1992) peut s'expliquer par les spécificités des systèmes orthographiques de chaque langue. Étant donné la divergence des résultats pour les différentes langues, ces auteurs insistent sur la nécessité d'approfondir la nature et le déroulement temporel des processus impliqués dans les dispositifs expérimentaux utilisés pour l'étude de ces phénomènes psycholinguistiques.

2.2.2 Mehler, Dupoux, & Segui (1990)

À partir de cette hypothèse syllabique, confortée par ailleurs par des résultats chez le très jeune enfant (voir Bertoncini, 1991), Mehler *et al.*, (1990) ont proposé le modèle SARAH. L'originalité de ce modèle repose sur une forte correspondance entre les processus d'accès au lexique chez l'adulte et les processus d'acquisition du langage chez l'enfant. L'unité syllabique est à la base de la constitu-

tion de représentations lexicales chez le jeune enfant, ce qui implique donc un rôle déterminant dans les processus d'accès au lexique à la fin du processus de maturation. Il s'agit d'un modèle *coarse-grained* dans lequel le traitement de l'information linguistique fait intervenir une unité de traitement de nature pré-lexicale, relativement large, apparentée à la syllabe. À la différence du modèle COHORT, le flux d'information entre les niveaux de traitement acoustico-phonétique et les niveaux supérieurs n'est pas continu ; il repose sur un niveau pré-lexical de traitement intermédiaire qui est responsable de l'accumulation d'informations acoustiques avant leur transmission aux niveaux supérieurs.

SARAH postule que la perception de la parole chez l'adulte repose sur trois niveaux de traitement : syllabique, lexical et phonologique. D'abord, une banque d'*analyseurs syllabiques* prend en charge la reconnaissance du *cadre syllabique.* Celui-ci correspond à une réalisation élémentaire, l'unité fonctionnelle minimale impliquée dans la production de la parole. Cette unité capture l'invariance au delà des différences de timbre, de débit d'élocution, d'accentuation et d'interlocuteur. L'inventaire des cadres syllabiques pour une langue donnée est assez réduit ; environ 6 000 pour le français. Au niveau du traitement lexical, la banque d'analyseurs syllabiques permet l'accès au lexique mental. Le processus est déclenché lorsqu'une quantité critique d'information est déjà traitée : la première syllabe d'un item constituant le code d'accès. Celle-ci représente la quantité minimale d'information susceptible d'activer une cohorte de candidats lexicaux. Enfin, les phonèmes sont dérivés des cadres syllabiques : dans SARAH, les phonèmes ne jouent pas un rôle direct dans la perception de la parole. Les auteurs soulignent que ce type de modèle explique un grand nombre de résultats obtenus dans des études psycholinguistiques. Ils reconnaissent, toutefois, que la nature exacte des cadres syllabiques reste à spécifier : cette unité ne correspond pas exactement à la syllabe telle qu'elle est définie dans les théories phonologiques et il est probable que ses caractéristiques varient selon des contraintes spécifiques pour chaque langue.

Comment l'inventaire des cadres syllabiques est-il constitué ? L'acquisition du lexique par le jeune enfant est rendue possible grâce à trois mécanismes de base : le filtrage syllabique, l'analyse phonétique et la détection de frontière de mot. Le *filtrage syllabique* est responsable du découpage du signal acoustique en segments syllabiques élémentaires. Ce filtre admet uniquement les unités syllabiques ayant des séquences acoustiques « légales » (il reste à expliquer les contraintes) ; les différences interlocuteur et le débit d'élocution ne sont pas prises en compte. SARAH propose un *analyseur phonétique* grâce auquel le très jeune enfant extrait la représentation phonétique sous-jacente à partir de représentations syllabiques. Celles-ci se présentent sous la forme d'un code abstrait qui fait intervenir les gestes nécessaires à sa production. La *détection de frontière de mot* utilise les représentations syllabiques et d'autres informations acoustiques (comme la durée et l'accent) pour calculer des indices élémentaires qui indiquent le début et la fin des mots.

Dans SARAH, le processus de maturation implique le passage d'un système phonétique à un système phonologique. La transition fait appel à deux mécanismes : la stabilisation sélective (ou désapprentissage) et la compilation. Le très jeune enfant possède un système universel de traitement de la parole lui permettant de discriminer tous les contrastes linguistiques observés dans les langues du monde. Avec la *stabilisation sélective* cette capacité universelle de traitement phonétique est réduite à un nombre restreint de contrastes phonétiques, ceux de l'ensemble de contrastes spécifiques de la langue maternelle. Cette spécialisation ne dépend pas de l'acquisition préalable d'un lexique mental, mais plutôt de l'acquisition de mécanismes d'extraction statistique et de fixation de paramètres. La *compilation,* présuppose un mécanisme d'acquisition supplémentaire destiné au stockage de gabarits syllabiques dans une mémoire à long terme,

permettant ainsi l'acquisition des représentations lexicales. À son tour, cet ensemble de mots potentiels active des niveaux plus élevés qui facilitent la stabilisation de la morphologie de la langue maternelle.

La plupart des modèles de la reconnaissance des mots ont été élaborés sur la base des performances des adultes. Le modèle SARAH propose une perspective de recherche différente basée sur un ensemble de contraintes qui interviennent dans la constitution des représentations lexicales au cours des premières années de la vie. Mehler *et al.*, (1990) affirment que la prise en compte des processus d'acquisition du langage permet de mettre en évidence les contraintes qu'il faut introduire dans la modélisation du processus de reconnaissance chez l'adulte. L'automaticité du processus d'acquisition du langage, la spécificité du système biologique (l'appareil phonatoire) et l'insensibilité aux variations de réalisation (par exemple, le débit d'élocution) placent l'acquisition de la parole au même niveau que les acquisitions spécialisées pour d'autres espèces animales : elle relève d'un apprentissage instinctif, de la fixation de paramètres et de la vérification d'hypothèses. Ainsi, les structures utilisées par l'adulte dépendent du bagage biologique de l'espèce et elles sont vraisemblablement liées aux structures de traitement de l'information présentes déjà dans son enfance. Au lieu d'être envisagée comme inhérente à des mécanismes résultant uniquement de l'influence de l'environnement linguistique, l'automaticité des processus de traitement de la parole chez l'adulte, en apparence « mystérieuse », pourrait donc être mieux comprise si elle était considérée comme la stabilisation du processus de traitement qui se met en place au cours de la maturation.

À l'heure actuelle, le modèle SARAH n'a pas été véritablement mis à l'épreuve. Les recherches récentes se sont limitées à valider l'hypothèse syllabique dans différentes langues avec variation de quelques paramètres dans la tâche de détection de phonèmes (comme les caractéristiques de la cible ou la relation entre celle-ci et le mot porteur). Ce modèle prometteur nécessite encore un approfondissement et une meilleure spécification des concepts pour pouvoir être opérationnel.

3. Conclusion et perspectives

Parmi les modèles de la reconnaissance des mots, nous avons retenu COHORT et SARAH qui proposent des approches originales intégrant les problèmes liés au traitement du signal de parole et à la représentation des mots dans le lexique mental. De plus, ces deux modèles ont été élaborés sur la base de recherches menées sur plusieurs langues. COHORT postule un système de reconnaissance basé sur une unité de traitement et de représentation semblable aux traits phonologiques, alors que SARAH met l'accent sur la syllabe. Le premier décrit un processus de traitement continu dans lequel l'information sensorielle est directement transmise aux niveaux supérieurs ; le deuxième postule un processus discontinu dans lequel intervient un niveau intermédiaire de traitement responsable de l'accumulation d'informations acoustiques avant le déclenchement du processus d'accès au lexique. Pour tester les deux modèles, des paradigmes expérimentaux ont été développés au cours des années ; parmi ceux-ci technique du fenêtrage et la tâche de détection de phonèmes.

COHORT et SARAH sont en réalité des modèles relativement complémentaires : nous les avons choisis pour illustrer les problématiques inhérentes à l'étude de la reconnaissance des mots, domaine qui présente, nous l'avons vu, de sérieux problèmes théoriques et méthodologiques. En effet, la plupart des modèles ne peuvent, ni prévoir l'effet simultané de plusieurs variables indépendantes sur la performance dans une situation de reconnaissance des mots, ni expliquer les

différences de ces effets en fonction de la tâche expérimentale. Sont envisagées actuellement deux stratégies complémentaires pour saisir la complexité des processus de reconnaissance de mots : la restriction du paradigme expérimental à certains aspects des modèles et leur implantation informatique.

Un autre problème, lui aussi très difficile à résoudre, est lié au manque de bases de données lexicales satisfaisantes pour tester et contrôler des biais. En français, par exemple, les bases de données sur la fréquence lexicale posent problème. Le *Trésor de la langue française* (*TLF*, 1971), qui a été constitué à partir de textes variés sur plusieurs siècles, est le plus utilisé. La fréquence d'usage des mots écrits est, pour les mots du vocabulaire courant, très différente de celle des mots parlés. Par exemple les mots *sel* et *peigne* courants dans la conversation quotidienne, ont une fréquence assez faible dans la langue écrite. À ce problème s'ajoute celui de l'actualité des données fournies ; la fréquence d'usage des mots en 1996 est assez différente de celle de 1896 !

Aux difficultés méthodologiques associées aux paradigmes expérimentaux s'ajoutent des problèmes théoriques fondamentaux comme celui de la segmentation et de la nature de l'unité minimale de traitement dans le processus de reconnaissance. L'avancement de la technologie peut jouer un rôle important dans le développement des dispositifs expérimentaux et dans la simulation des processus cognitifs. Il reste à explorer la reconnaissance des mots parlés en situation bimodale qui n'a pas été abordée jusqu'à présent.

Ce manque de recherches sur la reconnaissance des mots en présentation audiovisuelle peut sembler paradoxal. L'acte de parole – sauf dans certaines situations, comme celle du téléphone – est naturellement bimodal et les travaux sur la perception audiovisuelle de la parole ont bien montré que l'apport des informations visuelles des gestes faciaux impliqués dans la production des sons, les informations sur les mouvements labiaux notamment, ont une fonction non-négligeable. Sumby & Pollack (1954), et par la suite d'autres chercheurs comme Benoît, Mohamadi et Kandel pour le français, ont mis en évidence et chiffré l'apport que les gestes labiaux fournissent pour rendre la parole compréhensible lorsqu'elle est transmise en milieu bruité. Enfin, la vision peut, dans certains cas, biaiser l'interprétation des informations sonores non ambiguës. Les informations visuelles peuvent, en une situation de conflit, avoir une influence négative, allant jusqu'à produire des illusions perceptives : un stimulus auditif [ba] présenté simultanément avec un stimulus visuel [ga], est perçu comme [da] (McGurk & McDonald, 1976).

Les travaux sur la reconnaissance des mots ont fait d'énormes progrès dans les vingt dernières années, mais il reste encore beaucoup de recherches en perspective.

Références

Archangeli, D. (1984). *Underspecification in Yawelmani phonology and morphology.* Doctral dissertation, M.I.T., Cambridge.

Benoît, C., Mohamadi, T., & Kandel, S. (sous presse). Effect of phonetic context on audio-visual intelligibility of French. *Journal of Speech and Hearing Research*, 37, 1195-1203.

Bertoncini, J. (1991). Percevoir la parole sans les mots. In R. Kolinsky, J. Morais, & J. Segui (Eds.), *La reconnaissance des mots dans les différentes modalités sensorielles : Etudes de psycholinguistique cognitive* (p. 37-58). Paris : PUF.

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

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

Dupoux, E. (1989). *Identification des mots parlés. Détection de phonèmes et unité pré-lexicale.* Thèse doctorale, École de Hautes Études en Sciences Sociales, Grenoble II.

Dupoux, E., & Mehler, J. (1992). Unifying awareness and on-line studies of speech : A tentative framework. In J. Alegria, D. Holender, J. Morais, & M. Radeau (Eds.), *Analytic approaches to human cognition* (p. 59-75). North Holland : Elsevier.

Frauenfelder, U.H. (1991). Une introduction aux modèles de reconnaissance des mots parlés. In R. Kolinsky, J. Morais, & J. Segui (Eds.), *La reconnaissance des mots dans les différentes modalités sensorielles : Etudes de psycholinguistique cognitive* (p. 7-36). Paris : PUF.

Frauenfelder, U.H. (1992). The interface between acoustic-phonetic and lexical processing. In M.E.H. Schouten (Ed.), *The auditory processing of speech : From sound to words*). Berlin, New York : Mouton de Gruyter.

Frauenfelder, U.H. Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing : Facilitatory or inhibitory ? *Journal of Experimental Psychology : Human Perception and Performance, 16* (1), 77-91.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28* (4), 267-283.

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

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

Lahiri, A., & Marslen-Wilson, W.D. (1992). Lexical processing and phonological representation. In D.R. Ladd, & G.J. Docherty (Eds.), *Papers in laboratory phonology II, Gesture, segment, prosody* (p. 229-254). Cambridge : Cambridge University Press.

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

Liberman, A.M., Shankweiler, D., Fisher F.W., & Carter, B. (1974). Reading and the awareness of phonetic segments. *Journal of Experimental Child Psychology, 18*, 201-212.

Marcus, S.M., & Frauenfelder, U.H. (1985). Word recognition – uniqueness or deviation ? A theoretical note. *Language and Cognitive Processes, 1*, 163-169.

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

Marslen-Wilson, W.D. (1984). Function and process in spoken word recognition. In H. Bouma, & D.G. Bouwhuis (Eds.), *Attention and performance X : Control of language processes.* Hillsdale, N.J. : Lawrence Erlbaum Associates.

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

Marslen-Wilson, W.D. (1990). Activation, competition, and frequency in lexical access. In G. Altman (Ed.), *Cognitive models of speech processing : Psycholinguistic and computational perspectives* (p. 148-172). Cambridge : MIT Press.

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

McGurk, H., & McDonald, J. (1976). Hearing lips and voices. *Nature, 264*, 746-748.

McQueen, J. (1995). Processing versus representation : Comments on Ohala & Ohala. In B. Connell, A. Arvaniti, & I. Watson (Eds.), *Papers in laboratory phonology IV*, (pp. 61-67), Cambridge : Cambridge University Press.

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

Mehler, J., Dupoux, E., & Segui, J. (1990). Constraining models of lexical acces : The onset of word recognition. In G. Altman (Ed.), *Cognitive models of speech processing : Psycholinguistic and computational perspectives* (pp. 263-280). Cambridge : MIT Press.

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

Morais, J., Castro, S.L., & Kolinsky, R. (1991). La reconnaissance des mots chez les illettrés. In R. Kolinsky, J. Morais, & J. Segui (Eds.), *La reconnaissance des mots dans les différentes modalités sensorielles : Etudes de psycholinguistique cognitive* (p. 59-80). Paris : PUF.

Ohala, J., & Ohala, M. (1995). Speech perception and lexical representation : The role of vowel nasalization in Hindi and English. In B. Connell, A. Arvaniti, & I. Watson (Eds.), *Papers in laboratory phonology IV*, (pp. 41-60), Cambridge : Cambridge University Press.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (sous presse). Mora or syllable ? Speech segmentation in Japanese. *Journal of Memory and Language.*

Rietveld, T., & Frauenfelder, U. (1987). The effect of syllable structure on vowel duration. *Proceedings of the 11th International Congress of Phonetic Sciences*, Tallin.

Sebastian-Galles, N., Dupoux, E., Segui, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language, 31*, 18-32.

Segui, J. (1991). La reconnaissance visuelle des mots. In R. Kolinsky, J. Morais, & J. Segui (Eds.), *La reconnaissance des mots dans les différentes modalités sensorielles : Etudes de psycholinguistique cognitive* (p. 99-118). Paris : PUF.

Segui, J., Frauenfelder, U., & Mehler J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology, 72*, 471-477.

Sumby, W.H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212-215.

Warren, P., & Marslen-Wilson, W.D. (1987). Continuous uptake of acoustic cues in spoken word-recognition. *Perception and Psychophysics, 41* (3), 262-275.

Warren, P., & Marslen-Wilson, W.D. (1988). Cues to lexical choice : Discriminating place and voice. *Perception and Psychophysics, 43* (1), 21-30.

Zwisterlood, P., Schiefers, H., Lahiri, A., & Van Donselaar, W. (1991). *On the role of the syllable in the perception of Dutch.* Unpublished manuscript.