

La reconnaissance des mots parlés

Uli H. Frauenfelder

Noël Nguyen

Laboratoire de psycholinguistique
Faculté de Psychologie et des Sciences de l'Éducation
Université de Genève, Suisse

in: J.A. Rondal & X. Seron, eds. *Troubles du langage: Bases Théoriques, Diagnostic et Rééducation* (Mardaga, Bruxelles), pp. 213–240.

1 Introduction

Écouter et comprendre ce qui nous est dit: voilà une tâche que nous réalisons en permanence de manière automatique. Pourtant, pour le psycholinguiste qui se fixe pour but d'en étudier le fonctionnement, le traitement du langage oral soulève de multiples problèmes d'une grande complexité. En situation de communication ordinaire, l'auditeur est placé dans la nécessité de traiter en moyenne 200 mots par minute. Il ne dispose ainsi que de 300 millisecondes environ pour localiser chaque mot à l'intérieur d'un lexique mental contenant probablement entre 50000 et 100000 entrées. Dans la plupart des cas, ces échanges verbaux s'accomplissent dans un environnement bruyé. L'auditeur doit en outre faire face à l'extraordinaire variabilité présentée par les sons de la parole d'un locuteur à l'autre, et pour un même locuteur d'une situation de communication à l'autre. Il doit par-dessus tout se montrer capable d'établir une relation entre deux univers foncièrement hétérogènes, l'univers physique des sons de la parole, et l'univers symbolique des mots, des phrases et du discours. L'objet de cette contribution est de présenter en résumé ce que nous savons aujourd'hui sur les mécanismes cognitifs permettant à l'auditeur de s'acquitter de sa tâche avec autant d'efficacité.

Dans ce qui suit, nous traiterons essentiellement des processus employés par l'auditeur pour convertir le signal de parole en une séquence de mots, dans la mesure où ce qui fait suite à cette étape de traitement (analyse syntaxique / sémantique) n'est probablement pas spécifique à la modalité orale (voir Content, ce volume). La reconnaissance des mots se prête en outre à être étudiée de manière séparée, parce qu'elle se trouve selon toute vraisemblance assurée par un module de traitement (semi-)indépendant des autres. Pour déterminer le sens d'un énoncé, on s'accorde à penser en effet qu'il est nécessaire de passer par l'intermédiaire d'un **lexique mental**, dans lequel sont spécifiées de manière ad hoc les associations entre formes sonores et significations pour tous les mots connus de l'auditeur, ces associations revêtant comme on le sait un caractère arbitraire. La reconnaissance des mots repose ainsi probablement sur un ensemble de processus spécifiques aussi bien qu'essentiels à la compréhension du langage oral.

Sur la figure 1 sont présentés les principaux niveaux de traitement dans la reconnaissance des mots, tels que ces niveaux sont identifiés dans un modèle classique. À la base de ce schéma se trouve le signal de parole, que le système auditif transforme en ce que nous appellerons ici une **représentation d'entrée**. Il est supposé que cette représentation est découpée en segments (opération de **segmentation**) mis en relation chacun avec une unité phonétique (opération de **catégorisation**). De cela résulte une **représentation infra-lexicale**, qui se définit ainsi comme une séquence d'unités discrètes. La représentation infra-lexicale sert à ce titre-là d'interface, ou de zone de contact, entre le signal de parole et le lexique, dans la mesure où le signal, par essence continu, est ainsi converti sous une forme facilitant sa mise en relation avec les différentes entrées du lexique. On considère également que la représentation infra-lexicale sert à faire abstraction de tout ce qui, dans le signal, n'est pas essentiel à l'identification du mot prononcé (ex.: différences entre locuteurs), en allégeant ainsi la tâche du système de traitement.

— Insérer la figure 1 ici —

Le ou les mots encodés dans cette représentation infra-lexicale demandent à être alignés correctement avec les différentes entrées du lexique (opération d'**alignement**), puis à leur être comparés (opération d'**appariement**) de façon à déterminer quelle est l'entrée lexicale correcte pour chacun de ces mots. Nous appellerons **identification lexicale** le processus permettant d'aboutir de la représentation infra-lexicale à l'entrée lexicale correspondante. Le terme d'**accès au lexique** sera employé ici pour désigner le processus permettant à l'auditeur de prendre connaissance des différentes informations relatives à la forme (morpho/phonologique, orthographique) et au contenu (sémantique, syntaxique) d'une entrée lexicale (Frauenfelder, 1991). À chaque entrée se trouve en particulier associée une représentation se rapportant à sa forme sonore, et qui sera désignée sous le terme de **représentation phonologique lexicale**. Précisons enfin que le terme de **reconnaissance des mots** recouvrira dans ce texte à la fois l'identification lexicale et l'accès au lexique.

Le plan de ce chapitre est le suivant. Nous commencerons par décrire les principales propriétés

spécifiques à la modalité orale et les problèmes qui en découlent dans la perception de la parole et la reconnaissance des mots parlés (section 2). Puis, nous donnerons un aperçu des méthodes employées par les psycholinguistes dans ce domaine (section 3), pour présenter ensuite une typologie des modèles actuels de la reconnaissance des mots (section 4). Dans les sections suivantes (5–7), nous aborderons de manière plus détaillée les différents niveaux de traitement définis dans notre modèle de base, en montrant en quoi les données expérimentales dont nous disposons à ce jour nous permettent de donner davantage de contenu à ce modèle.

2 Les défis de l'entrée parlée

Le signal de parole présente plusieurs propriétés qui lui sont spécifiques et ne se rencontrent pas dans l'écriture, et qui posent un véritable défi aux chercheurs s'efforçant de comprendre les processus sous-jacents au traitement lexical. En premier lieu, la parole est un phénomène *directionnel*. Elle est étalée dans le temps, et possède par définition un début, un milieu et une fin. De ce fait, l'auditeur reçoit les informations relatives au mot-cible bout par bout seulement. Les processus mis en œuvre dans la reconnaissance de ce mot sont ainsi assujettis à une contrainte temporelle externe: l'ordre dans lequel les sons de la parole aboutissent à l'oreille (Mattys, 1997). Cette contrainte ne s'applique pas dans la lecture dans la mesure où toute l'information nécessaire à l'identification d'un mot écrit est immédiatement accessible à l'œil (pour autant que ce mot ne dépasse pas une certaine longueur).

En deuxième lieu, la parole est *continue*. Contrairement à l'écriture, le signal de parole ne comporte pas d'“espaces”, ou de périodes de silence signalant à l'auditeur où se situent les frontières entre phonèmes par exemple ou entre mots. Le caractère continu, ininterrompu de la parole soulève un problème majeur qui est celui du passage entre continu et discret, c'est-à-dire la mise en correspondance entre un signal d'entrée continu et des représentations lexicales discrètes.

En troisième lieu, la parole est *variable*: un mot n'est jamais produit deux fois exactement de la même façon, et il présente des différences substantielles sur le plan phonologique et/ou phonétique selon le locuteur (son âge, son sexe, son origine) par exemple. Chaque mot se matérialise ainsi par une infinité de formes sonores différentes, que l'auditeur doit se montrer capable de ramener à une entité lexicale unique. En outre, les processus phonologiques mis en application dans certains contextes phonologiques, ainsi que le débit de parole, contribuent à modifier davantage encore la forme sonore de chaque mot. Cette variabilité fait de la relation entre formes sonores et entités lexicales une relation complexe (non-biunivoque).

La variabilité des sons de la parole est en partie attribuée aux phénomènes dits de **coarticulation**, que nous aurons l'occasion de mentionner à plusieurs reprises. On désigne par coarticulation le fait que les mouvements accomplis par les articulateurs dans la production de la parole se chevauchent sur l'axe temporel (Hardcastle & Hewlett, sous presse). Dans une syllabe de type CV par exemple, les gestes articulatoires associés à la consonne initiale et à la voyelle qui la suit sont partiellement superposés. Il en résulte que chaque portion du signal est le plus souvent à mettre en relation avec plusieurs unités phonétiques à la fois dans la chaîne parlée, et que chaque unité phonétique se matérialise par des indices acoustiques distribués en différents points de ce signal.

Ces trois propriétés – directionnalité, continuité et variabilité — compliquent la tâche du système de reconnaissance des mots, dans la forme qui lui est donnée sur la figure 1 du moins. Pour que les mots puissent être identifiés correctement, il est nécessaire en fait que soient résolus deux problèmes majeurs, le problème de la **segmentation**, et celui de la **catégorisation**.

Le problème de la segmentation est lié au fait que le signal de parole se laisse difficilement découper en portions associées chacune à une unité linguistique et une seule, qu'il s'agisse de phonèmes, de syllabes, ou de mots. Ce problème se pose d'abord dans le passage entre signal de parole et représentation infra-lexicale. Si l'on postule que cette représentation prend la forme d'une chaîne de

phonèmes par exemple, il est extrêmement difficile de repérer dans le signal des événements acoustiques susceptibles de coïncider de manière systématique avec des frontières entre phonèmes. Le problème de la segmentation se présente également dans la mise en œuvre de cette opération d'alignement entre la représentation infra-lexicale et les différentes entrées du lexique. La question pour l'auditeur est alors de déterminer où commence et se termine chaque mot, en l'absence de pauses entre les mots. La difficulté de la tâche tient en partie au fait que le signal parvient à l'auditeur linéairement (directionnalité), et qu'une séquence de phonèmes interprétable comme un mot peut se révéler faire partie en fait d'un mot plus long (problème des mots enchâssés dans d'autres mots).

Le problème de la catégorisation trouve son origine dans le fait que les sons de la parole présentent une variabilité extrêmement large imputable à de multiples sources (variabilité inter-locuteurs, variabilité contextuelle, etc.; cf. Perkell & Klatt, 1986). Ce problème se pose lui aussi – en des termes différents – à chaque niveau de traitement. Au niveau infra-lexical en premier lieu, la correspondance entre formes sonores et unités infra-lexicales possède comme nous l'avons vu un caractère non-biunivoque. Lorsque l'on cherche ainsi à partitionner l'espace vocalique en un ensemble de régions correspondant chacune à une voyelle, dans une langue donnée, ces régions sont le plus souvent marquées par des recouvrements partiels (Peterson & Barney, 1952). En deuxième lieu, la relation entre représentation infra-lexicale et lexique se montre elle-même non-biunivoque: un même mot peut être associé à différentes représentations infra-lexicales (ex.: "quatre" peut se prononcer [katR] ou, dans un style plus familier, [kat]); à l'inverse, des mots de signification différente peuvent se prononcer de la même manière (homophones). Par suite, l'auditeur se heurte également à un problème de catégorisation quand il lui faut mettre en correspondance la représentation infra-lexicale avec les entrées du lexique (processus d'appariement).

Le problème de la segmentation et celui de la catégorisation peuvent trouver différentes solutions dans le cadre du modèle classique présenté sur la figure 1, mais ils ont également amené certains psycholinguistes à remettre en question certaines des hypothèses de base de ce modèle, comme

nous le verrons plus tard. Dans la section suivante, nous passons en revue quelques méthodes employées dans les recherches sur la reconnaissance des mots parlés.

3 Méthodes d'étude

Les psycholinguistes sont de plus en plus nombreux à aborder le traitement lexical selon une approche nouvelle, consistant à combiner données expérimentales, simulations réalisées au moyen de modèles computationnels, et données quantitatives sur le lexique établies à partir de bases de données informatisées. La figure 2 illustre la manière dont les trois sources d'information sont combinées dans cette approche.

— Insérer la figure 2 ici —

3.1 Techniques expérimentales temps réel

Ces techniques consistent à demander aux sujets de répondre aussi rapidement que possible (c.à.d. en quelques centièmes de milliseconde) aux stimuli qui leur sont présentés, et à mesurer leur temps de réponse pour chaque stimulus. Les temps de réponse (ci-après TR) et les pourcentages d'erreur permettent alors de procéder à des inférences sur les caractéristiques du traitement mis en jeu. Les techniques chronométriques constituent la méthode la plus fréquemment employée en psycholinguistique expérimentale, particulièrement dans le champ de la reconnaissance des mots (cf. Grosjean & Frauenfelder, 1997, pour une présentation des différentes techniques utilisées). En recourant à des tâches simples (détection de cible, décision binaire, répétition immédiate), et en tentant de réduire l'intervalle temporel séparant la présentation du stimulus et la réponse produite par le sujet, on peut espérer que les données recueillies reflètent avec une certaine fidélité la nature et la durée des opérations mentales mises en œuvre.

3.2 Modèles computationnels

Les modèles computationnels se fondent sur des programmes informatiques visant à simuler les mécanismes de traitement mis en œuvre par l'humain. Ces modèles nous offrent le moyen de rendre compte de phénomènes hautement complexes (cf. Dijkstra & de Smedt, 1996). Leur emploi présente plusieurs avantages importants (Content & Frauenfelder, 1996). D'une part, ils contraignent le modélisateur à définir chaque processus de traitement avec une grande précision, ce qui fait souvent défaut dans les théories présentées sous une forme purement verbale. En outre, on peut en analysant son comportement s'assurer que le modèle est complet et cohérent, et tester ainsi de manière préliminaire sa plausibilité. Enfin, les modèles de simulation donnent lieu à des prédictions quantitatives, beaucoup plus précises en tant que telles que les prédictions faites par les modèles "verbaux", et qui peuvent être directement comparées avec des données expérimentales (cf. figure 2). Dans le domaine de la reconnaissance des mots parlés, plusieurs modèles de simulation ont été proposés, et ont eu une influence considérable sur l'évolution des idées et des recherches entreprises. Ce fut le cas en particulier de TRACE (McClelland & Elman, 1986), que nous présenterons plus loin (4.5).

3.3 Statistiques lexicales

Le développement des technologies de la langue a un impact sensible en psycholinguistique, notamment à travers l'utilisation accrue de bases de données lexicales informatisées. Ces bases de données comportent des informations de différente nature (informations phonologiques, orthographiques, morphologiques, syntaxiques et sémantiques, fréquences d'utilisation, etc.) sur les mots contenus dans une langue. Des bases de données lexicales existent à présent pour un certain nombre de langues. Citons ainsi CELEX (voir le site Web www.kun.nl/celex ainsi que Burnage, 1990) pour l'anglais, l'allemand et le néerlandais, ainsi que BRULEX (Content, Mousty & Radeau, 1990) pour le français. Ces bases de données donnent lieu à des analyses statistiques permettant de caractériser de manière quantitative les propriétés structurales d'une langue à différents niveaux d'analyse. Elles offrent également la possibilité de procéder à des comparaisons dans ce domaine entre des

langues différentes (cf. par exemple Goldman, Content & Frauenfelder, 1996). Les bases de données lexicales remplissent en outre deux autres fonctions plus pragmatiques, mais qui n'en sont pas moins essentielles pour l'expérimentation et la simulation. Elles permettent en premier lieu de sélectionner des stimuli pour les besoins d'une expérience en contrôlant aussi soigneusement que possible les variables parasites potentielles (Cutler, 1981). En deuxième lieu, c'est à partir de ces bases de données que peuvent être construits les mini-lexiques utilisés dans les études de simulation.

Ces trois sources d'information – données expérimentales, simulations sur ordinateur et statistiques lexicales – peuvent donc être combinées et comparées de différentes manières. On peut leur ajouter aujourd'hui des données d'une autre nature encore, qui nous sont fournies par la neuropsychologie, et qui revêtent à l'évidence aussi une importance majeure. Les études réalisées sur des patients cérébrolésés ont apporté des indications fondamentales sur le rapport entre traitement lexical et cerveau. Les recherches faisant appel à l'imagerie cérébrale (Kutas & van Petten, 1994) offrent aujourd'hui l'espoir de mieux identifier les structures corticales impliquées dans le traitement de la parole, ainsi que de caractériser avec plus de précision le déroulement temporel de ce traitement.

Dans la section suivante sont présentés avec plus de détails les modèles computationnels développés aujourd'hui dans le but de simuler les processus cognitifs mis en jeu dans la reconnaissance des mots parlés.

4 Modèles de la reconnaissance des mots parlés

De multiples modèles ont été proposés pour rendre compte de la reconnaissance de mots, et ces modèles présentent une grande variété. Nombre d'entre eux sont des modèles de type "verbal", c'est-à-dire définis en des termes empruntés au langage ordinaire. C'est le cas du modèle Cohort (Marslen-Wilson & Welsh, 1978), entre autres exemples. Ces modèles verbaux sont à différencier des modèles de type computationnel, lesquels revêtent un caractère beaucoup plus formel puisqu'ils

se présentent sous la forme d'un programme informatique. Au sein des modèles computationnels, on peut en outre distinguer les modèles de type traitement de l'information, basés sur une opposition classique entre processus de traitement et représentations (ex.: modèle FLMP, Massaro, 1998), et les modèles connexionnistes (réseaux de neurones artificiels), introduits plus récemment. Au-delà de ces grandes catégories, on peut établir un certain nombre de distinctions supplémentaires se rapportant plus directement à la reconnaissance des mots.

4.1 Mécanismes d'appariement lexical: activation directe ou recherche sérielle?

Les modèles de la reconnaissance des mots peuvent en premier lieu être classifiés en fonction du nombre de comparaisons entre la représentation infra-lexicale et les différentes représentations lexicales que chaque modèle suppose pouvoir s'accomplir en même temps. Selon les modèles d'**activation directe** (cf. Marslen-Wilson & Welsh, 1978; Morton, 1969), la représentation infra-lexicale est comparée simultanément avec toutes les entrées lexicales. L'état de chaque entrée ou son **niveau d'activation** évolue en fonction de son degré de correspondance avec le signal de parole. À l'inverse, dans les modèles de **recherche lexicale sérielle** (cf. Bradley & Forster, 1987; Forster, 1976) – les entrées lexicales sont examinées l'une après l'autre. L'identification lexicale consiste en une exploration sérielle à travers une liste d'entrées lexicales. Le temps nécessaire pour identifier une entrée lexicale est supposé dépendre du nombre de comparaisons à effectuer de manière successive avant que l'entrée appropriée soit enfin rencontrée.

La prédominance des modèles d'activation directe pour la reconnaissance des mots parlés peut pour une large part être attribuée aux propriétés du signal de parole. La parole est étalée dans le temps, et de ce fait l'auditeur ne reçoit les informations relatives au mot-cible que bout par bout. Par suite, il n'est pas évident de savoir à quel moment, selon un modèle de recherche sérielle, l'auditeur est supposé mettre ou remettre en route une recherche lexicale (cette recherche doit-elle par exemple débiter après le premier, le deuxième ou le troisième segment?). Ce problème ne se

pose pas pour des mots écrits.

4.2 Modèles localistes et modèles distribués

La majeure partie des modèles de la reconnaissance des mots parlés (Cohort, TRACE, SHORTLIST) se rangent dans la classe des modèles **localistes**. Ces modèles reposent sur l'idée que chaque entrée lexicale est représentée par une unité dont le niveau d'activation est proportionnel au degré de correspondance de cette unité avec le signal. On considère qu'un mot a été reconnu lorsque son niveau d'activation dépasse un certain seuil (ou satisfait à un autre critère de même nature). D'autres modèles connexionnistes, basés sur des représentations **distribuées**, ont également été proposés dans le domaine de la reconnaissance des mots parlés (Gaskell, Hare & Marslen-Wilson, 1995). Dans les modèles de ce type, les représentations lexicales revêtent une forme distribuée. Ainsi, chaque mot est représenté par un ensemble d'unités, et une même unité peut réciproquement être associée à différents mots. Nous nous limiterons ici à discuter de modèles localistes, en raison de leur homogénéité, de leur valeur heuristique, et de leur usage très répandu.

4.3 Modèles autonomes et modèles interactifs

Les modèles psycholinguistiques peuvent également être distingués selon la manière dont ils définissent les effets de contexte, et selon qu'ils postulent ou non que le traitement de la parole fait intervenir un flux d'information de haut en bas. Dans les modèles **autonomes**, les processus de bas en haut ne sont pas modifiés par le contexte et donc aboutissent à la reconnaissance lexicale sans tenir compte de l'information des niveaux supérieurs. On supposera par exemple que la phrase dans laquelle apparaît un mot ne peut avoir d'influence sur les processus permettant à ce mot d'être reconnu (Forster, 1979). A l'opposé, dans les modèles **interactifs** (Marslen-Wilson & Tyler, 1980) l'information contextuelle peut exercer un effet sur le traitement de bas en haut à différents niveaux. Dans le modèle interactif TRACE par exemple, le niveau d'activation d'un phonème est déterminé par les informations provenant à la fois du niveau inférieur (détecteurs de trait) et du niveau lexical.

4.4 Flux d'activation et sélection lexicale

On s'accorde généralement à penser que l'identification lexicale se fonde sur l'activation d'un ensemble de compétiteurs lexicaux, et sur la sélection du mot-cible dans cet ensemble. Selon toute vraisemblance, c'est un flux d'information de bas en haut, partant de la représentation infra-lexicale et aboutissant au lexique, qui permet à un ensemble de candidats de se mettre en place.

En revanche, le consensus est moins grand en ce qui concerne la manière dont les candidats sont éliminés de l'ensemble des compétiteurs. On peut établir une distinction entre deux principaux mécanismes de sélection. Selon le premier, la sélection lexicale s'accomplit au moyen d'une **inhibition de bas en haut**. Lorsque l'information sensorielle reçue cesse d'être compatible avec un candidat, ce candidat est désactivé. Selon le second mécanisme, la réduction du nombre de compétiteurs s'opère à travers un processus **d'inhibition latérale**. Cette inhibition entre compétiteurs lexicaux permet à ceux dont le niveau d'activation est le plus élevé, et en particulier au mot-cible, de prédominer et d'éliminer les compétiteurs plus faibles. Ces deux mécanismes ne sont pas mutuellement exclusifs et peuvent être combinés dans le même modèle.

4.5 Un exemple: le modèle TRACE

Parmi les différents modèles que nous venons de mentionner, TRACE (McClelland & Elman, 1986) est un exemple bien adapté au cadre de cette revue des travaux en raison des multiples discussions qu'il a suscitées. C'est un modèle de la reconnaissance des mots de type connexionniste et localiste. Il se compose d'un grand nombre d'unités de traitement connectées les unes aux autres à l'image des réseaux de neurones dans le cerveau.

Les unités de traitement se répartissent dans ce modèle 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. Elles se caractérisent par un certain niveau d'activation, proportionnel à leur degré de correspondance avec les informations qui leur sont envoyées.

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). En outre, des connexions inhibitrices sont établies latéralement entre unités de même niveau (trait-trait, phonème-phonème et mot-mot).

Le système de traitement est mis en route lorsque le signal de parole vient activer la couche des traits. L'unité relative au trait "voisé" par exemple, réagira à la présence de voisement dans le signal. Les détecteurs de traits activent à leur tour les unités phonémiques qui leur sont associées (ainsi, le trait "voisé" exercera un effet activateur sur tous les phonèmes voisés, /b/, /z/, /m/, etc.). De la même manière, les phonèmes dont le niveau d'activation dépasse le seuil de repos activent les mots qui les contiennent.

L'activation se propage dans le réseau de bas en haut, du niveau des traits jusqu'à celui des mots, mais aussi de haut en bas, du niveau des mots vers celui des phonèmes. Il est ainsi supposé que chaque mot contribue à accroître le niveau d'activation des phonèmes dont il se compose. Par ailleurs, en raison de la présence de connexions inhibitrices latérales, l'augmentation du niveau d'activation d'une unité de traitement s'accomplit au détriment des autres unités de même niveau. C'est grâce à ce mécanisme qu'une unité fortement activée peut réduire à zéro l'influence des unités moins activées.

La mise en relation entre signal de parole et lexique ne s'accomplit pas instantanément dans ce modèle. On suppose que l'information se diffuse progressivement d'un niveau à l'autre. Le système est en fait placé sous le contrôle d'une sorte d'horloge interne, qui régit la vitesse avec laquelle l'information se propage. Le processus s'accomplit de manière itérative, pas par pas, sous la forme d'une séquence de cycles de traitement.

Il est également important de noter TRACE est un modèle *parallèle*, au sens où toutes les unités de traitement entrent en fonctionnement dès que le signal de parole aboutit au réseau. Cela signifie en particulier que les détecteurs de mots sont activés bien avant que soient identifiés tous les phonèmes dont le mot-cible est composé.

En résumé, TRACE est un modèle localiste, interactif, et appartenant à la famille “activation directe”. Dans ce modèle, il est supposé que le signal de parole est analysé sous la forme d’un ensemble de traits distinctifs. Ces traits sont mis en relation avec le lexique par l’intermédiaire d’une représentation infra-lexicale de type phonémique. La sélection du mot-cible parmi l’ensemble des compétiteurs se fonde sur un mécanisme d’inhibition latérale. Le niveau lexical exerce un effet activateur de type top-down sur celui des phonèmes.

5 La représentation infra-lexicale

On s’est beaucoup interrogé sur la structure interne de la représentation infra-lexicale que l’on suppose être construite par l’auditeur dans l’identification lexicale. Dans la plupart des modèles actuels du traitement de la parole, ces représentations sont décomposables sous la forme d’une séquence d’unités élémentaires, le plus souvent définies en termes linguistiques: le phonème, la syllabe, et le trait, en particulier. Rappelons brièvement que l’on désigne par **phonème** une unité distinctive minimale (impossible à décomposer en une succession de segments plus petits à valeur distinctive), par **syllabe** un groupe phonémique constitué d’un phonème appelé noyau (une voyelle le plus souvent) et, facultativement, d’une attaque et/ou d’une coda, et par **trait** une dimension phonétique servant à opposer deux séries de phonèmes (ex.: voisé/non-voisé, continu/interrompu, etc.).

Comme nous allons le voir, les unités dont se compose la représentation infra-lexicale ont alternativement été assimilées à des phonèmes, à des syllabes ou à des traits. Dans cette section,

nous passons brièvement en revue les données expérimentales recueillies dans le but de mettre à l'épreuve chaque hypothèse, en commençant par l'hypothèse phonémique.

5.1 Le phonème comme unité de représentation infra-lexicale

Il a longtemps été supposé que la représentation infra-lexicale consistait en une séquence linéaire de phonèmes (voir par exemple Marslen-Wilson & Welsh, 1978; Pisoni & Luce, 1987). Cette idée trouve en partie son origine dans les théories phonologiques construites autour de la notion de phonème, dont l'influence a été grande sur les premières recherches menées en psycholinguistique. Le phonème offre également des avantages en ce qui concerne le stockage des entrées lexicales, chaque entrée étant munie dans cette hypothèse d'une forme construite à partir d'un nombre minimal d'unités de base (environ 35 phonèmes en français), ce qui permettrait de réduire la place occupée par le lexique dans la mémoire à long terme. (Notons cependant que les problèmes du stockage ne doivent pas être confondus avec les problèmes de traitement.)

Depuis le début des années 1950, les mécanismes mis en œuvre dans l'identification des phonèmes ont fait l'objet d'une multitude de travaux en phonétique. Les recherches réalisées aux laboratoires Haskins (Liberman, 1996) en particulier, ont fait prévaloir une hypothèse fondamentale en vertu de laquelle les phonèmes sont perçus sur un mode **catégoriel** (Harnad, 1987). Lorsqu'il est demandé à des sujets d'identifier des sons prenant place sur un continuum entre deux extrêmes clairement reconnaissables (ex.: /p/-/b/), les réponses obtenues basculent brutalement au milieu du continuum entre la première et la deuxième catégorie. Les sujets se montrent en outre mieux capables de discriminer deux sons lorsque ces derniers sont perçus comme étant associés à des phonèmes différents plutôt qu'au même phonème, toutes choses égales d'ailleurs. On peut interpréter ce phénomène en disant que l'auditeur est peu sensible aux différences entre sons rattachés à une même catégorie phonématique. La notion de perception catégorielle a cependant suscité différentes critiques formulées entre autres par Massaro (Massaro & Cohen, 1983), qui voit en elle une simple forme de réponse induite par la tâche soumise au sujet (choix binaire). La théorie des

aimants perceptifs (perceptual magnets) proposée plus récemment par Kuhl (1991) réintroduit la notion de perception catégorielle sous une forme affaiblie, à travers l'idée que les sons sont identifiés par comparaison avec des "prototypes", et que les différences perçues entre sons s'amenuisent progressivement au voisinage de chaque prototype.

Cependant, soulignons dès à présent que le rôle du phonème dans le traitement de la parole demande encore à être clairement établi. Les expériences venant d'être citées font apparaître que l'auditeur se montre capable d'identifier des phonèmes lorsque cela lui est demandé, mais elles ne permettent pas d'affirmer que la reconnaissance des mots s'opère à partir d'une représentation infra-lexicale de type phonémique. Par ailleurs, les expériences faisant appel à des techniques temps réel (ex.: détection de fragment) donnent généralement à observer une primauté de la syllabe sur le phonème (voir section 5.2).

En outre, les modèles phonémiques se doivent d'apporter une solution à ces deux problèmes majeurs que nous avons appelés problème de la segmentation et problème de la catégorisation. Ces problèmes dérivent pour une part du moins des phénomènes de coarticulation, dont nous avons souligné la prévalence dans la production de la parole (section 2). Les effets de coarticulation ne permettent pas que le signal de parole puisse être découpé en une suite de morceaux séparés par des frontières clairement repérables. Ils donnent en outre à penser que chaque phonème est soumis à l'influence des phonèmes adjacents, en étant ainsi produit sous une forme différente d'un contexte à l'autre. Les solutions apportées à ces problèmes se laissent ranger en deux grandes catégories.

En premier lieu, on a supposé que la variabilité des sons de la parole, plutôt que de constituer un bruit rendant l'identification des phonèmes plus difficile, forme en fait une *source d'information* mise à profit en tant que telle par l'auditeur (Elman & McClelland, 1988). Les phénomènes de coarticulation par exemple sont assujettis à des lois que l'on commence à bien connaître (Hardcastle & Hewlett, sous presse). Les variations présentées par un phonème sous l'influence du contexte

revêtent en d'autres termes un caractère systématique et régulier. Il est souvent postulé à présent que l'auditeur utilise ces régularités à son profit en se rapportant d'une manière ou d'une autre au contexte pour identifier chaque phonème. Cette hypothèse est implémentée sous une forme numérique dans le modèle TRACE, entre autres exemples.

En second lieu, il est possible de remettre directement en question le postulat selon lequel le signal de parole doit être décomposé par l'auditeur en une séquence *linéaire* de segments, chaque segment débutant là où le précédent se termine. Comme nous l'avons indiqué, les phénomènes de coarticulation confèrent en fait au signal une structure non-linéaire caractérisée par le fait que les segments se chevauchent partiellement sur l'axe temporel. Dans le modèle d'analyse vectorielle perceptive de Fowler (1984), on suppose que l'auditeur se représente le signal à l'image de la manière dont celui-ci est produit, c'est-à-dire sous la forme d'une séquence de segments partiellement superposés. Selon Fowler, cette représentation non-linéaire permet à l'auditeur de s'affranchir des problèmes de segmentation et de catégorisation (le lecteur est renvoyé à Fowler, 1984, pour plus de détails).

5.2 La syllabe comme unité de représentation infra-lexicale

Les psycholinguistes ont entrepris de résoudre le problème de la variabilité d'une autre manière encore, en remettant directement en question la thèse selon laquelle la reconnaissance d'un mot passe par l'identification de phonèmes. Selon Mehler (1981) par exemple, ce sont les syllabes qui constituent les unités perceptives de base dans le traitement de la parole. Cette hypothèse se fonde sur l'idée que les effets de coarticulation sont plus marqués à l'intérieur d'une syllabe qu'à la frontière entre deux syllabes. Les syllabes présenteraient ainsi moins de variations en fonction du contexte que les phonèmes, au sens où les syllabes résisteraient chacune davantage à l'influence des syllabes adjacentes que ne le feraient les phonèmes à celle des phonèmes adjacents. L'hypothèse syllabique trouve également son origine dans le fait que tout locuteur semble posséder une connaissance intuitive de la notion de syllabe (en se montrant capable de décompter ou de permuter des

syllabes dans un mot), alors que la notion de phonème ne semble faire surface à la conscience qu'avec l'apprentissage de la lecture (Morais, Cary, Alegria & Bertelson, 1979).

L'hypothèse selon laquelle la syllabe est l'unité perceptive de base dans le traitement de la parole a donné lieu à des investigations plus directes faisant appel à des tâches on-line, telles que la détection de fragments (Frauenfelder & Kearns, 1996). Dans ce type de tâche, les sujets se voient présenter une cible, consistant en un phonème ou en une séquence de phonèmes, et qui leur est spécifiée sous une forme visuelle (lettres) ou auditive. Cette cible est suivie d'un stimulus acoustique, la tâche des sujets étant alors de déterminer aussi rapidement que possible si la cible se trouve ou non contenue dans le stimulus. En comparant les TR obtenus selon que la cible coïncide avec le phonème initial ou avec la syllabe initiale dans le stimulus, on peut tenter d'établir laquelle de ces deux unités, phonème ou syllabe, prime sur l'autre dans le traitement de la parole. Les multiples expériences construites sur ce modèle depuis le début des années 1970, ont abouti à la conclusion qu'une syllabe-cible est détectée plus rapidement qu'un phonème-cible et donc constitue l'unité perceptive. Soulignons cependant que certains résultats expérimentaux dont nous disposons dans ce domaine ont montré l'effet inverse, en donnant à observer une primauté du phonème sur la syllabe (Norris & Cutler, 1988).

Le rôle de la syllabe dans le traitement de la parole a été établi pour le français dans une expérience réalisée par Mehler, Dommergues, Frauenfelder et Segui (1981). Dans cette expérience, les sujets avaient pour tâche de détecter aussi rapidement que possible une cible prédéterminée de type CV (ex.: BA), ou de type CVC (ex.: BAL) dans une séquence sonore disyllabique dont la syllabe initiale était également soit de type CV (ex.: "balance"), soit de type CVC (ex.: "balcon"). Les résultats montrèrent que la cible visuelle était détectée plus rapidement lorsqu'elle coïncidait avec la syllabe initiale du mot porteur, indépendamment de la longueur de cette cible (2 ou 3 phonèmes). La cible BAL par exemple donnait lieu à des TR plus courts que la cible BA dans le mot "balcon", et à des TR plus longs dans le mot "balance". De tels résultats sont en désaccord avec un modèle

phonémique de la perception de la parole, dans la mesure où celui-ci aboutirait à prédire que les cibles les plus courtes (CV) sont toujours détectées plus rapidement que les cibles les plus longues (CVC), qu'elles coïncident ou non avec la première syllabe de la séquence porteuse.

5.3 Le trait comme unité de représentation infra-lexicale

Les difficultés associées à la notion de segment (variabilité acoustique, absence de frontières entre chaque segment et le segment suivant) ont amené certains chercheurs à renoncer à penser que la reconnaissance des mots reposait sur la construction d'une représentation infra-lexicale segmentale, que les unités dont celle-ci serait formée soient de type phonémique ou de type syllabique. Pour Stevens (1986) et Marslen-Wilson et Warren (1994), entre autres, l'identification lexicale s'accomplit directement à partir d'une matrice de *traits asynchrones*. 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 (voir supra). Dans les modèles de Stevens et de Marslen-Wilson, un trait est mis en relation avec un mot par une voie directe, plutôt que de l'être par l'intermédiaire d'une unité phonémique. Dans la même perspective, plutôt que d'être assemblés en faisceaux correspondant chacun à un phonème, les traits sont ici à considérer comme évoluant dans le temps de manière (semi-)indépendante. Le problème de la segmentation cesse tout simplement de se poser, et celui de la variabilité contextuelle est reformulé en des termes nouveaux, puisqu'il est supposé que les phénomènes de coarticulation donnent essentiellement lieu à des modifications dans l'organisation temporelle des traits, chaque trait restant associé à un ensemble de corrélats acoustiques invariants.

L'hypothèse selon laquelle un lien direct est établi entre traits et lexique dans la reconnaissance des mots a fait l'objet d'une série d'expériences récentes (Warren & Marslen-Wilson, 1987, 1988; Lahiri & Marslen-Wilson, 1991) basées sur la méthode de présentation de stimuli auditifs dite du dévoilement graduel (gating, voir Grosjean, 1996). Cette méthode consiste à présenter un mot-cible par morceaux, ou portes, de durée croissante, les sujets ayant pour consigne de deviner après

chaque porte quel est ce mot. La technique permet ainsi d'établir avec précision à partir de quel point dans le signal il devient possible au sujet d'identifier correctement le mot présenté. Dans une expérience réalisée sur l'anglais britannique par exemple, Warren et Marslen-Wilson (1987) ont fait apparaître qu'un auditeur se montre 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. Selon Warren et Marslen-Wilson, ces résultats montrent que le trait de nasalité entre directement en jeu dans la sélection des unités lexicales dès que sa présence est détectée dans le signal, alors qu'un modèle phonémique classique conduirait à prédire que le mot est reconnu plus tardivement, à la présentation de la consonne finale.

Les modèles de l'identification lexicale à partir d'une représentation en traits laissent en suspens plusieurs questions. En premier lieu, si l'on suppose que les traits sont distribués de manière non-linéaire dans la représentation infra-lexicale, les règles présidant à cette organisation restent encore à définir. En pratique, lorsque l'on entreprend d'implémenter un modèle de ce type sous la forme d'un programme par exemple (cf. Gaskell, Hare & Marslen-Wilson, 1995), les traits restent disposés en colonnes associées chacune à un phonème, ce qui signifie que le modèle demeure implicitement segmental. Par ailleurs, les traits utilisés présentent le plus souvent un caractère abstrait (voisé, nasal...), et ils continuent en tant que tels de soulever le problème de l'invariance (en d'autres termes, la question d'identifier les corrélats acoustiques associés à chaque trait et à lui seul). Enfin, on peut penser que les patrons de réponse obtenus dans une expérience de gating demeurent compatibles avec les modèles segmentaux, dès lors qu'il est admis que les segments peuvent empiéter l'un sur l'autre sur l'axe temporel (Fowler, 1984; McClelland & Elman, 1986).

5.4 Problèmes en suspens et solutions possibles

Comme nous le voyons, les expériences visant à déterminer la nature des unités de représentation infra-lexicales dans le traitement de la parole ont abouti à des résultats disparates. Trois facteurs au moins ont été mis en avant pour expliquer le fait qu'il soit difficile d'identifier une unité de base.

En premier lieu, une distinction est à établir entre unités de segmentation d'une part, et unités de catégorisation d'autre part. Cutler et Norris (1988) ont ainsi suggéré que le processus de segmentation et le processus de catégorisation s'accomplissent de manière indépendante, et sur la base d'unités de taille différente. Selon ces auteurs, la segmentation se fonde en anglais sur les syllabes accentuées, tandis que la catégorisation s'applique à des unités dont la taille est inférieure à celle de la syllabe, peut-être des phonèmes.

En deuxième lieu, un grand nombre de travaux s'insérant dans une tradition comparative donnent à penser que représentations et unités sont susceptibles de varier dans leur structure d'une langue à l'autre. Ainsi, des recherches inter-langues faisant appel à la tâche de détection de fragment (tâche déjà décrite dans la section 5.2) ont abouti à des résultats assez variés en fonction de la langue examinée. Cutler et collaborateurs (Cutler et al., 1983) ont ainsi conduit une série d'expériences en français et en anglais, dont les résultats donnèrent à observer un effet syllabique en français seulement. D'autres études réalisées sur le japonais, le catalan, l'espagnol et le hollandais ont donné lieu à des résultats différents selon les propriétés phonologiques de chaque langue (Kolinsky, 1998).

En deuxième lieu, l'idée selon laquelle le traitement de la parole se fonde sur une unité de représentation infra-lexicale et une seule est un postulat que l'on peut également remettre en question. Certains modèles récents (voir par ex. Kolinsky, 1998) établissent une distinction entre plusieurs étapes de traitement (perceptuelles et post-perceptuelles) pouvant faire appel à des unités différentes. Il est souvent difficile pour le chercheur de déterminer avec précision à quel niveau de traitement doivent être rapportés les résultats observés, dans la mesure où la réponse dépend de la procédure expérimentale utilisée (détection de fragment par exemple, ou amorçage phonologique), et des conditions expérimentales.

Pour conclure, remarquons que nous avons peu discuté ici de la nature des informations employées

par l'auditeur dans la construction de la représentation infra-lexicale à partir du signal de parole (question des indices acoustiques, entre autres choses). De fait, le passage entre signal et représentation infra-lexicale est un problème laissé en suspens dans la majeure partie des modèles actuels de la reconnaissance des mots (à l'exception notable de TRACE I, qui se présentait comme un système complet de reconnaissance automatique permettant en théorie d'identifier une séquence de mots à partir d'un signal de parole naturel, voir McClelland & Elman, 1986). Pour la plupart d'entre eux, ces modèles prennent la représentation infra-lexicale pour point de départ, sans rendre compte des processus aboutissant à la mise en place de cette représentation. Cette absence s'explique en partie par 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. Les recherches les plus récentes dans le domaine de la reconnaissance des mots visent à abolir cette division.

Nous abordons à présent les processus pouvant permettre à l'auditeur de passer de la représentation infra-lexicale au lexique.

6 Des représentations infra-lexicales au lexique:

l'identification lexicale

Dans cette entreprise visant à caractériser les processus mis en jeu dans l'identification lexicale, tout modèle se doit de spécifier la manière dont la représentation infra-lexicale construite à partir du signal est alignée et comparée avec les représentations stockées dans le lexique mental. Ces opérations d'appariement et d'alignement sont discutées l'une après l'autre dans les deux prochaines sections.

6.1 L'appariement et l'activation lexicale

De nombreuses expériences ont été conduites dans le but de caractériser le processus permettant à un mot-cible d'être sélectionné au sein d'un ensemble initial de candidats lexicaux. Notre objectif n'est pas ici de passer en revue tous les résultats obtenus, mais plutôt d'illustrer la logique sous-jacente à ces expériences et de résumer leurs principales conclusions. Nous examinerons d'abord les données expérimentales relatives à l'activation par le signal des candidats appariés et alignés avec l'entrée (c.à.d. formant ce que l'on appelle la cohorte initiale). Nous présenterons ensuite les données expérimentales pouvant être citées en faveur de l'un ou de l'autre des deux mécanismes de sélection décrits plus haut (4.4).

6.1.1 L'activation du bas en haut

La manière dont les entrées du lexique sont activées par le signal de parole fait l'objet d'un ensemble de propositions très précises dans le modèle Cohort (Marslen-Wilson & Welsh, 1978). Selon ce modèle, sont activés dans un premier temps les mots dont la partie initiale (premier ou deux premiers phonèmes) correspond exactement avec celle de la représentation infra-lexicale. Ces mots constituent la cohorte initiale. Dans un second temps, et au fur et à mesure que se poursuit le traitement du signal d'entrée, les mots candidats sont éliminés les uns après les autres de la cohorte dès qu'ils cessent de correspondre avec le signal (désactivation causée par un mécanisme d'inhibition de bas en haut). Un mot présenté hors contexte est donc reconnu à partir du moment où il est le seul à figurer encore dans la cohorte. Ce moment est appelé le **point de reconnaissance**. Il est supposé coïncider dans Cohort avec le **point d'unicité** du mot, lequel désigne le phonème à partir duquel le mot devient unique dans le lexique. Ainsi, on postulera que le mot "éléphant" ([elefã]) est reconnu à partir du phonème /f/, dans la mesure où le lexique ne comporte pas d'autre mot commençant lui aussi par la séquence "éléph". En d'autres termes, Cohort établit ici une équivalence entre une variable psychologique, dont la valeur est établie de manière expérimentale, le point de reconnaissance, avec une variable structurale, le point d'unicité, dont la valeur peut être déterminée à partir d'une base de données lexicales.

Les recherches portant sur l'activation des candidats lexicaux font appel à différentes procédures expérimentales, comprenant la détection de phonèmes (Connine & Titone, 1996), la détection de mots enchâssés dans des non-mots (word spotting, McQueen, 1996), et l'amorçage sémantique transmodal (cross-modal semantic priming; Swinney, 1979; Tabossi, 1996). Dans cette procédure, on présente au sujet une phrase parlée, ou bien encore une liste de mots parlés isolés (ou de non-mots). Pendant ou juste après l'audition du mot-amorce, le sujet se voit présenter sur écran une séquence de lettres écrites, la tâche étant de déterminer si cette séquence est un mot ou un non-mot (décision lexicale). Dans la condition critique, le mot visuel possède une relation sémantique avec le mot présenté auditivement, ou avec l'un de ses compétiteurs. Les TR enregistrés dans cette condition sont comparés avec les TR obtenus dans une condition de contrôle (absence de lien sémantique entre amorce auditive et séquence visuelle). Toute réduction du TR dans la condition avec lien par rapport à la condition sans lien est interprétée comme indiquant que l'amorce ou ses compétiteurs ont donné lieu à une activation sémantique.

Dans une étude importante basée sur cette procédure, Zwitserlood (1989) a fait apparaître que le sens du mot-cible et celui du mot compétiteur sont tous les deux activés, aussi longtemps que l'information sensorielle ne permet pas de les différencier. Ainsi, lorsque le fragment de mot parlé [kapit] est présenté au sujet, le sens du mot "capitaine" et celui du mot "capital" sont simultanément activés, en facilitant une décision lexicale portant sur des mots qui leur sont reliés sémantiquement tels que "bateau" et "argent", respectivement, par rapport à une situation de contrôle neutre. Ce résultat est en accord avec les modèles avec activation de bas en haut, et avec l'idée selon laquelle tous les compétiteurs appariés et alignés avec le mot d'entrée sont activés au cours du traitement.

La manière dont les candidats lexicaux non-appariés sont traités par l'auditeur n'est pas encore clairement établie. La question se pose donc de savoir ce qui se produit lorsque le mot présenté à

l'auditeur est mal prononcé (ex.: “déléphone” au lieu de “téléphone”). Selon le modèle Cohort – dans sa version initiale du moins – toute erreur de prononciation située avant le point d'unicité devrait empêcher que le mot-cible soit activé, étant donné que signal et mot-cible ne sont pas parfaitement appariés en leur début. Une première étude (Marslen-Wilson & Zwitserlood, 1989) faisant appel à l'amorçage sémantique transmodal et réalisée sur le hollandais suggère que seuls les candidats lexicaux appariés sont activés. Les auteurs montrèrent qu'un mot ou un non-mot parlé, servant d'amorce, et dont le phonème initial présente avec celui du mot-cible des différences portant sur plusieurs traits distinctifs, n'active pas le sens de ce mot-cible (ainsi, ni le mot rime “mat” ni le non-mot rime “dat” n'activaient le mot-cible “cat”, et ils étaient par conséquent sans effet sur la décision lexicale visant son voisin sémantique “dog”). Connine, Blasko et Titone (1993) ont également employé la technique de l'amorçage transmodal, et ils ont manipulé la distance phonologique entre un non-mot servant d'amorce et le mot-cible. Les résultats donnèrent à observer un effet d'amorçage pour les non-mots créés chacun à partir d'un mot, moyennant une petite modification apportée à la partie initiale de ce mot (changement dans la valeur d'un trait distinctif, ex.: voisement). Cet effet d'amorçage disparaissait lorsque la distance phonologique entre amorce et cible était augmentée.

Nous pouvons tirer de ces études la conclusion suivante. Les candidats appariés dans leur partie initiale avec le signal d'entrée sont, comme on peut s'y attendre, activés par ce signal. Lorsque le mot-cible est soumis à une distorsion, il semble pouvoir donner lieu malgré tout à une certaine activation lexicale, mais à la condition que cette distorsion ne transforme pas ce mot-cible en un autre mot. Pour que cette activation ait lieu, il est en outre nécessaire que le mot prononcé ne présente avec le mot-cible que de petites différences phonologiques en début de mot.

6.1.2 Inhibition latérale et inhibition de bas en haut

Il semble donc acquis qu'un ensemble de mots est activé pendant le traitement lexical. La question qui est posée maintenant est de savoir de quelle manière l'auditeur parvient à éliminer les

candidats non appropriés. Nous avons évoqué deux mécanismes différents: l'inhibition latérale et l'inhibition de bas en haut.

Les effets d'inhibition lexicale ont donné lieu à peu de travaux expérimentaux, probablement parce qu'ils sont difficiles à établir. Une expérience de word spotting réalisée par McQueen, Norris et Cutler (1994) nous fournit quelques indications à ce sujet. Les sujets avaient pour tâche de détecter des mots (ex.: "mess") situés en position non-initiale soit dans une séquence de phonèmes formant le début d'un mot plus long (ex.: "domes" [dɔmes]), soit à l'intérieur d'un non-mot de contrôle (ex.: "nomes" [nɔmes]). La détection du mot enchâssé était retardée par le compétiteur qui venait le recouvrir (ex.: "domestic"). Les TR plus longs pour les mots enchâssés dans des fragments de mot donnent à penser qu'un mot porteur plus long entre en compétition, et inhibe, le mot enchâssé. Ces résultats s'accordent avec l'hypothèse d'une inhibition latérale entre mots.

Des données en faveur d'une inhibition de bas en haut ont été obtenues par Frauenfelder, Content et Scholten (en prép.). Dans cette étude, les sujets avaient pour consigne de détecter aussi rapidement que possible des phonèmes-cible prédéterminés. Les séquences porteuses étaient pour une partie d'entre elles des non-mots construits en modifiant un phonème en position non-initiale dans un mot de référence (ex.: "vocabunaire"). Le phonème à détecter se situait après le phonème modifié (ex.: /R/, situé après /n/ dans le non-mot précité). Les TR ne présentaient pas de différence significative avec des non-mots de contrôle non susceptibles de donner lieu à une activation lexicale (ex.: "satobunaire"), mais ils étaient beaucoup plus lents que les TR relatifs aux mots d'origine ("vocabulaire"). Ces résultats suggèrent qu'un phonème non-apparié (/n/ dans "vocabunaire") désactive immédiatement le candidat lexical compatible avec la première partie du stimulus. De telles données sont en accord avec les modèles comportant un mécanisme d'inhibition latérale entre mots.

En résumé, on peut citer aujourd'hui quelques données expérimentales allant dans le sens d'une inhibition de bas en haut comme dans celui d'une inhibition latérale. Ces deux mécanismes peuvent

bien sûr être combinés au sein du même modèle, comme c'est le cas dans le modèle **SHORTLIST** (Norris, 1994).

6.2 La segmentation et l'alignement lexical

Comme nous l'avons déjà souligné, le signal de parole est continu, et les frontières entre mots n'y apparaissent pas de manière systématique. Le problème qui se pose ainsi à l'auditeur est d'aligner correctement signal d'entrée et représentation lexicale. Par alignement, nous désignons le fait que l'auditeur doit déterminer quelle partie de la représentation d'entrée est à mettre en relation, ou à comparer, avec quelle partie des représentations lexicales. Les solutions proposées face à ce problème d'alignement peuvent être classifiées selon le niveau de traitement et le type d'information utilisé pour accomplir l'alignement. On peut de ce point de vue-là établir une distinction entre les stratégies de segmentation **infra-lexicales**, basées sur le traitement de l'information contenue dans le signal, et les stratégies de segmentation **lexicales**, fondées comme leur nom l'indique sur un traitement opéré au niveau lexical.

6.2.1 Stratégies de segmentation infra-lexicales

Les stratégies de segmentation infra-lexicales proposées font appel à des informations (acoustiques et prosodiques) de nature assez variée, depuis les indices allophoniques (Church, 1987), aux contraintes phonotactiques (Frazier, 1987) et à la structure prosodique (Cutler & Butterfield, 1992).

C'est à la fin des années 50 que l'on a commencé de chercher des indices phonétiques fiables associés aux frontières de mots. Les phonéticiens ont à cette époque entrepris de premières analyses acoustiques détaillées sur des séquences "quasi-homophones" (ex.: "grand tamis" vs. "grand ami"), de manière à identifier l'information discriminante. Différents indices spécifiques à la langue ont été mis en évidence en anglais (Lehiste, 1960) et en suédois (Garding, 1967), comprenant par exemple l'allongement de durée, l'insertion d'un coup de glotte, etc. Du point de vue perceptif, Nakatani

et Dukes (1977) ont montré que l'auditeur est capable de segmenter correctement des séquences ambiguës (telles que “no notion” et “known ocean”) à partir d'indices phonétiques. Cependant, l'utilité pour la perception de ces indices phonétiques de frontière apparaît relativement limitée, dans la mesure où ils ne sont pas systématiquement présents dans le signal.

Les contraintes phonotactiques gouvernent l'ordre dans lequel les phonèmes peuvent se succéder à l'intérieur d'un mot, au sein de la même syllabe comme à la frontière entre deux syllabes. Elles constituent une source d'information qui semble pouvoir servir de support à la segmentation lexicale. En français, certaines séquences de consonnes sont illégales en début ou en fin de syllabe (ex.: /pz/), alors qu'elles peuvent se rencontrer à la frontière entre deux mots. Ces séquences semblent faciliter la segmentation du signal de parole en mots (Banel & Bacri, 1997). Des résultats analogues ont été obtenus pour le hollandais (McQueen & Cox, 1995) et le finnois (Suomi, McQueen & Cutler, 1997), bien que la nature des contraintes en question soit bien entendu différente d'une langue à l'autre.

Récemment, l'attention des chercheurs s'est davantage portée sur les informations de nature prosodique. Il a été suggéré par Cutler et Norris (1988) que les auditeurs de langue maternelle anglaise appliquaient une stratégie de segmentation basée sur les régularités propres à la prosodie de l'anglais. Selon leur stratégie de segmentation métrique (SSM), l'auditeur postule qu'une frontière de mot prend place immédiatement avant chaque syllabe forte (ne contenant pas de voyelle réduite). Chaque syllabe de ce type déclenche ainsi une recherche lexicale. Cette heuristique aboutit généralement à un résultat correct dans la mesure où la plupart des mots anglais à contenu débutent par une syllabe forte (Cutler & Carter, 1987). Les résultats expérimentaux présentés par Cutler et ses collègues suggèrent que les auditeurs anglophones segmentent bien le signal de parole selon cette stratégie. Dans une expérience importante (Cutler & Norris, 1988), les sujets avaient pour tâche de détecter des mots monosyllabiques (ex.: “mint”) enchâssés en position initiale dans un mot disyllabique dont la seconde syllabe était accentuée (ex.: /mIn'telf/) ou non-accentuée (ex.:

/^lmIntəf/). Les TR se montrèrent plus longs dans le premier cas, un effet que les auteurs attribuèrent au fait que deux recherches lexicales sont simultanément mises en route (/mIn/ et /təIf/), le /t/ étant perçu comme formant le début de la seconde syllabe plutôt la fin du mot “mint”. En français, la structure métrique (alternance de syllabes longues et courtes) semble également fournir des indices pour la segmentation (Banel & Bacri, 1997). Les sujets avaient à décider si des séquences ambiguës (ex.: “bordure”) correspondaient à un ou deux mots. Lorsque le patron métrique était de type bref-long, les sujets répondaient que la séquence constituait un mot simple, alors que le patron long-bref donnait lieu à la réponse opposée (deux mots). Ces résultats s’accordent avec l’hypothèse selon laquelle l’information prosodique joue un rôle important dans la segmentation lexicale, la manière dont cette information est utilisée variant toutefois en fonction de la structure prosodique de la langue.

6.2.2 Stratégies de segmentation lexicales

On distingue également différents types de segmentation lexicale selon le type de traitement lexical impliqué. Dans le modèle TRACE par exemple (McClelland & Elman, 1986), la segmentation se fonde sur une compétition lexicale entre les candidats activés. L’item lexical dont l’activation est la plus forte inhibe et désactive, à travers un mécanisme d’inhibition latérale, les items alignés avec le signal à partir d’un point différent. L’item remportant la compétition constitue le mot reconnu, et l’emplacement de ses frontières avec le mot précédent et le mot suivant devient du même coup connu de l’auditeur. Ainsi la segmentation fait suite à l’identification lexicale, et elle en constitue l’un des résultats. Selon une autre hypothèse, proposée dans le modèle Cohort (Marslen-Wilson & Welsh, 1978), la segmentation lexicale s’appuie sur le produit de l’accès au lexique. L’auditeur est ici supposé accéder à la représentation phonologique associée à un mot avant d’avoir atteint la fin de ce mot. Une fois que cette information phonologique lui devient accessible, l’auditeur a la possibilité de prédire la fin du mot en cours de traitement et le début du mot suivant.

Ces différentes stratégies de segmentation infra-lexicales et lexicales ne sont pas mutuellement

exclusives. La segmentation peut résulter d'un traitement faisant intervenir de multiples indices à des niveaux différents. Ainsi, la procédure de segmentation implémentée dans la version modifiée du modèle `SHORTLIST` (Norris, McQueen & Cutler, 1995) s'appuie à la fois sur un mécanisme d'inhibition latérale et sur des indices prosodiques (la SSM).

7 Les représentations lexicales

Dans cette dernière section, nous étudierons de plus près la représentation se tenant au point d'arrivée dans notre modèle de base (fig. 1), c'est-à-dire la représentation lexicale. Deux grandes questions se posent au psycholinguiste dans ce domaine: 1) Quelle est la structure interne de chaque entrée lexicale? 2) De quelle manière les entrées sont-elles organisées à l'intérieur du lexique? Dans ce qui suit, nous aborderons les représentations lexicales sous l'angle phonologique, puis sous l'angle morphologique.

7.1 Aspects phonologiques

Comme nous l'avons déjà indiqué, il est supposé ici qu'à chaque entrée lexicale est associée une forme phonologique spécifiant la manière dont cette entrée se prononce. Cette représentation phonologique lexicale est caractérisée en des termes très différents d'un modèle à l'autre. On peut ranger dans une première catégorie les modèles réduisant les entrées lexicales à de simples **étiquettes** sans forme phonologique propre. Dans le modèle `TRACE` par exemple, chaque mot se rattache à une unité de traitement dépourvue de structure interne. La forme phonologique de ce mot se matérialise en fait à travers les liens qui s'établissent entre ce mot et les détecteurs de phonème situés sur la couche inférieure. Dans le modèle proposé par Lahiri et Marslen-Wilson (1991) en revanche, les entrées lexicales sont de type **structuré**, dans la mesure où chaque entrée se trouve dotée d'une forme phonologique déterminée de manière explicite.

Les modèles à représentations lexicales structurées se répartissent eux-mêmes sur un continuum en fonction du degré d'abstraction de ces représentations. À une extrémité du continuum se situe

la théorie de la sous-spécification phonologique (Archangeli, 1988), telle qu'elle est appliquée à la reconnaissance des mots parlés par Lahiri et Marslen-Wilson (1991). Selon ces auteurs, la représentation phonologique rattachée à chaque entrée dans le lexique est **abstraite** et sous-spécifiée, au sens où cette représentation est formée par un nombre minimal de traits distinctifs, traits dont la valeur respective est spécifiée si et seulement si elle se différencie de celle qui leur est attribuée par défaut. Lahiri et Marslen-Wilson postulent que c'est en se référant à cette représentation abstraite sous-spécifiée, plutôt qu'à une représentation phonétique de surface, que l'auditeur interprète le signal de parole dans la reconnaissance des mots. À l'autre extrémité du continuum vient prendre place le modèle défendu par Pisoni et son groupe (Palmeri et al., 1993; Pisoni, 1993). Contrairement à la théorie de la sous-spécification, ces travaux donnent à penser que les représentations phonologiques stockées dans le lexique sont extrêmement **concrètes**, en contenant par exemple des informations propres à chacun des locuteurs connus de l'auditeur. Ainsi, lorsque des sujets ont pour tâche de dire si le mot qui leur est présenté est "ancien" (déjà présenté antérieurement) ou "nouveau", il a été montré que les TR sont plus courts quand les mots ont été prononcés par le même locuteur plutôt que par des locuteurs différents.

7.2 Aspects morphologiques

La place de la morphologie dans l'organisation du lexique, comme en ce qui concerne la structure de chaque entrée lexicale, a fait l'objet de nombreuses recherches, essentiellement centrées sur la modalité visuelle. Pour ce qui touche à la structure interne des entrées du lexique, les hypothèses proposées prennent place sur un continuum, dont l'hypothèse décompositionnelle (Taft & Forster, 1976), et l'hypothèse du listing exhaustif (Butterworth, 1983), constituent les deux extrêmes. Selon la première hypothèse, les mots complexes sont codés dans le lexique sous une forme décomposée avec une représentation séparée pour chaque morphème et des règles de composition permettant de les combiner. À l'inverse, les modèles de type listing exhaustif se fondent sur un postulat selon lequel les mots morphologiquement complexes sont listés sous une forme unitaire. Les modèles "hybrides" (Caramazza, Laudanna & Romani, 1988), à mi-chemin entre ces deux extrêmes, se

montrent plus nuancés. Ils supposent une double représentation des mots (décomposée et non-décomposée) et ils font entrer en jeu différents facteurs tels que le statut lexical (mot/non-mot), la fréquence d'utilisation des mots et des morphèmes, les types de morphème (préfixe/suffixe et inflexionnel/dérivationnel), ou encore la langue. Nous renvoyons le lecteur intéressé à la revue des travaux récemment publiée sur le sujet par McQueen & Cutler (1997).

8 Effets sur le traitement lexical

La vitesse et la précision avec lesquelles l'auditeur identifie un mot obéissent à différents facteurs. Ces facteurs se rapportent aux propriétés structurales (composition, longueur phonologique et morphologique) et distributionnelles des mots (leur fréquence et celle des différentes unités - phonèmes, diphtongues, syllabes et morphèmes - dont ils se composent). En outre, le contexte phrastique dans lequel les mots se trouvent insérés - en d'autres termes, leur prédictabilité - peut également affecter les processus de reconnaissance. Une discussion de l'ensemble de ces facteurs dépasserait le cadre de ce travail. Nous nous pencherons ici sur l'influence exercée par la fréquence d'utilisation d'un mot sur la reconnaissance de ce mot. Nous examinerons également le rôle du contexte phrastique et lexical dans le traitement de la parole.

8.1 Effet de fréquence

Plus un mot est fréquent dans une langue, plus vite et mieux il est reconnu: c'est ce que les psycholinguistes désignent par l'**effet de fréquence**. Bien qu'il s'agisse là de l'un des effets les plus fiables jamais établis dans le domaine du traitement du langage (oral comme écrit), on continue de chercher à mieux cerner son origine. Dans les premières expériences réalisées sur ce point (Savin, 1963) les sujets avaient à identifier des mots présentés dans du bruit. Les résultats montrèrent que les mots fréquents étaient mieux reconnus que les mots rares. Des expériences faisant appel à des tâches de décision lexicale (Tyler, Marslen-Wilson, Rentoul & Hannay, 1988) et de détection de phonème (Dupoux & Mehler, 1990), ont également permis de montrer qu'un mot donne lieu à un TR plus court lorsqu'il est plus fréquent.

Différentes explications ont été avancées pour rendre compte de l'effet de la fréquence lexicale. Certains chercheurs (Morton, 1969; voir également le modèle TRACE) considèrent que l'effet de fréquence se manifeste très tôt dans le traitement, les mots plus fréquents se caractérisant par un niveau d'activation de base plus élevé que les mots plus rares. Selon d'autres chercheurs (Luce et al., 1990), cet effet se produit à une étape plus tardive, faisant suite à l'activation initiale des candidats lexicaux. Selon ces auteurs, l'effet de fréquence intervient lors de la décision "post-accès" permettant de déterminer quel est le mot retenu dans la sélection lexicale. À l'heure actuelle, les données empiriques dont nous disposons ne permettent pas de trancher de manière définitive entre ces différentes hypothèses, mais elles penchent plutôt dans le sens d'un effet tardif (Connine, Titone & Wang, 1993).

8.2 Les effets de contexte

Comme nous l'avons vu plus haut (4.3) les modèles de la reconnaissance lexicale ne s'accordent pas sur le rôle attribué au contexte dans le traitement lexical. Il est utile de distinguer deux types de contexte ici: le contexte lexical et le contexte phrastique. Nous traiterons ainsi de l'influence possible des représentations lexicales sur le traitement infra-lexical, en premier lieu, et de celle du contexte phrastique sur l'identification lexicale, en deuxième lieu.

8.2.1 Les effets lexicaux

De manière générale, on parle d'**effet lexical** lorsqu'un son de parole est interprété différemment par l'auditeur selon le statut lexical de la séquence porteuse (mot/non-mot). Les effets lexicaux sont au centre du débat entre modèles autonomes et modèles interactifs (section 4.3), et ils ont à ce titre suscité de multiples travaux depuis le début des années 1980 (voir Pitt & Samuel, 1993, pour une revue des travaux).

Les effets lexicaux se manifestent sous différentes formes selon la tâche expérimentale utilisée.

L'une des premières expériences réalisées sur le sujet (Warren, 1970) a consisté à présenter à des auditeurs une série de mots dont un phonème avait été préalablement remplacé par du bruit (ex.: /s/ dans "legislatures"). Les auditeurs avaient pour consigne d'indiquer si, selon eux, le bruit venait se substituer au phonème-cible, ou s'il avait été simplement superposé à ce phonème. La majeure partie des sujets percevaient le signal comme étant intact (bruit superposé), en conduisant ainsi Warren à conclure que le phonème manquant avait donné lieu à un processus de **restauration** perceptive, sous l'influence du contexte lexical. Dans une expérience ultérieure marquée par différentes améliorations sur le plan méthodologique, Samuel (1981) a fait apparaître que le lexique semblait exercer un effet sur la manière même dont le signal était perçu, plutôt que de se limiter à biaiser la réponse du sujet dans un sens ou dans un autre à un niveau post-perceptif. Selon Samuel donc, le phénomène de restauration phonémique résulte d'un véritable transfert d'information de type top-down, et il s'accorde en cela avec les modèles interactifs du traitement de la parole.

Il a également été montré que le lexique exerce une influence sur la façon dont les sons de la parole sont interprétés dans une tâche d'identification de phonème. Dans l'expérience réalisée par Ganong en 1980 et devenue célèbre depuis, les sujets avaient pour tâche d'identifier une occlusive sur un continuum entre un mot (ex.: "dash") et un non-mot ("tash"). Les résultats ont montré que les sujets optaient plus fréquemment pour la réponse formant un mot avec la séquence porteuse (dans l'exemple cité, "d"), la tendance étant plus forte pour les stimuli les plus ambigus (au milieu du continuum). Cet effet a été interprété par Ganong comme allant également dans le sens des modèles interactifs. Dans une expérience construite sur le même modèle cependant, Fox (1984) a observé que l'influence du lexique se manifestait davantage pour les réponses lentes que pour les réponses rapides. Selon l'auteur, ces résultats faisaient apparaître que le lexique entre en jeu sous la forme d'un biais (en faveur de la réponse formant un mot) *après* l'identification du phonème-cible, et n'exerce donc pas d'influence directe sur les processus sous-jacents à cette identification.

À sa publication, l'expérience conduite par Elman et McClelland (1988) a été considérée par beau-

coup comme donnant à l'approche interactive un avantage majeur sur les modèles autonomes. 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. Ils se sont plus précisément proposé de déterminer si une consonne fricative ambiguë, mais dont le lexique permet de rétablir l'identité ((ex.: "ChristmaS", "fooliS", S représentant une fricative à mi-chemin entre /s/ et /ʃ/), peut avoir un effet sur la manière dont est identifiée une consonne occlusive adjacente (ex. "?ape", ? représentant une occlusive à mi-chemin entre /d/ et /g/). Les réponses observées présentaient les variations attendues en fonction du mot précédent. 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. Dans un travail plus récent cependant, Norris (1992) a démontré que l'effet mis en évidence par Elman et McClelland pouvait être simulé par un modèle connexionniste purement bottom-up.

Les effets lexicaux continuent de donner lieu à de multiples recherches visant à opposer modèles autonomes et modèles interactifs. Le lecteur est renvoyé à Norris, McQueen & Cutler (soumis pour publication), pour une récente synthèse sur le sujet.

8.2.2 Les effets phrastiques

Selon les théories autonomes, les informations syntaxiques et sémantiques fournies par le contexte phrastique dans lequel le mot se présente n'ont pas d'influence sur les processus mis en jeu dans la reconnaissance de ce mot. Le contexte n'intervient qu'à une étape ultérieure, au cours de laquelle l'auditeur procède à l'évaluation et à l'intégration des informations dont il dispose une fois que le mot a été identifié. Les modèles interactifs, en revanche, laissent supposer que l'information contextuelle de niveau phrastique contribue directement à l'identification lexicale. Dans certains modèles interactifs (Morton, 1969) par exemple, les attentes générées chez l'auditeur par

le contexte qui précède le mot à reconnaître donnent lieu à l'activation d'un certain nombre de candidats lexicaux qui ne correspondent pas toujours au signal d'entrée. D'autres modèles interactifs (Marslen-Wilson, 1984) attribuent au contexte un rôle moins fort en postulant qu'il est utilisé pour éliminer des candidats lexicaux déjà activés par le signal.

De nombreuses études ont mis en évidence un effet facilitateur du contexte. Par exemple, des expériences de détection de cible lexicale (Marslen-Wilson & Tyler, 1980) ont montré que les mots-cible sont détectés plus rapidement dans des phrases grammaticalement correctes que dans des séquences agrammaticales ou sémantiquement anormales. Cependant, si ces résultats montrent clairement que le contexte peut exercer un effet facilitateur, ils ne permettent pas de déterminer la nature exacte de celui-ci, qui peut s'expliquer soit par une augmentation du niveau d'activation du mot-cible, soit par une intégration plus facile de ce mot-cible dans l'interprétation de la phrase. Les données expérimentales recueillies dans ce domaine ne permettent pas de trancher en faveur de l'une ou l'autre de ces deux hypothèses.

Les études portant sur la reconnaissance des mots homophones ont abouti à des données plus claires. Dans les expériences de ce type (voir par ex. Swinney, 1979), les sujets entendent un mot interprétable de deux manières différentes (ex.: "maire/mère") à l'intérieur d'une phrase excluant l'une de ces deux significations. Lorsque l'on compare, au moyen de la tâche d'amorçage transmodal, les niveaux d'activation respectifs des deux candidats lexicaux (approprié/inapproprié) juste après la présentation de ce mot ambigu, on constate que ces niveaux d'activation sont équivalents. En revanche, des mesures effectuées quelques centaines de millisecondes après la fin du mot montrent que seul le candidat approprié est encore activé à ce moment-là. Lorsque le mot présenté est ambigu sur le plan syntaxique (ex.: "montre", substantif dans "la montre", verbe dans "je montre"), il a également été constaté que les deux interprétations possibles présentaient un niveau d'activation analogue (Tanenhaus, Leiman & Seidenberg, 1979) malgré le fait que le contexte supprimait théoriquement cette ambiguïté. De tels résultats suggèrent que le contexte

sémantique/syntaxique ne désactive pas d'emblée les interprétations inappropriées d'un mot et les mots incompatibles avec ce contexte (voir Zwitserlood, 1989, discuté dans 6.1.1). Ces études vont dans le sens d'un modèle autonome du traitement de la parole.

9 Conclusion

Dans ce chapitre, nous avons passé en revue quelques questions importantes qui se posent dans les recherches sur le traitement du langage oral. Nous avons pris pour point de départ un modèle simple comportant deux modules de traitement principaux. Le premier module a pour fonction de convertir le signal de parole, variable et continu, en une représentation infra-lexicale. Le deuxième sert à identifier l'entrée appropriée à l'intérieur du lexique mental à partir de cette représentation. Nous avons également présenté différentes données empiriques qui nous ont permis de donner davantage de substance à ce modèle et de le rendre plus précis sur de nombreux points. Au terme de ce tour d'horizon, nous espérons avoir fait apparaître les progrès considérables réalisés ces dernières années par les chercheurs dans ce domaine.

Nombre de questions majeures restent néanmoins à résoudre. Les efforts investis dans la mise à l'épreuve des modèles de la reconnaissance des mots se heurtent encore à de nombreux problèmes méthodologiques. Face à la multitude de variables indépendantes (qualité du stimulus, fréquence lexicale, longueur des mots, point d'unicité, contexte, etc.), les variables dépendantes sont au contraire en nombre réduit. Les méthodes dont nous disposons pour recueillir des informations sur le traitement lexical sont trop peu nombreuses encore. L'utilisation des techniques expérimentales de type temps réel ont fourni aux psycholinguistes la capacité d'étudier le déroulement temporel de la reconnaissance des mots avec une précision accrue. Il est fort probable que les techniques d'imagerie cérébrale apporteront une contribution nouvelle dans ce domaine en nous fournissant des données convergentes.

Dans les années à venir, l'un des principaux challenges pour les psycholinguistes s'attachant à étudier la reconnaissance des mots parlés – aussi bien que pour les ingénieurs cherchant à développer des systèmes automatiques de reconnaissance de la parole – sera de traiter le problème de la reconnaissance des mots dans la parole continue spontanée. Jusqu'à aujourd'hui, la plupart des travaux sur la reconnaissance des mots ont été réalisés avec des mots isolés, souvent articulés de manière soignée, plus rarement avec de la parole lue continue.

Un autre challenge consistera à mettre en relation les processus de traitement que l'on suppose être employés chez l'adulte, avec les mécanismes mis en œuvre dans l'acquisition du langage chez le bébé. Dans le cas du traitement du langage oral, cette contrainte a été considérée avec grand sérieux. En s'employant à mettre en place un système de reconnaissance des mots comparable à celui de l'adulte, les bébés ont à résoudre un problème non trivial qui est de découvrir les mots appartenant à leur langue maternelle sans avoir de connaissances préalables à ce sujet.

Remerciements

Cet article a été rédigé avec le support financier du Fonds National pour la Recherche Scientifique suisse (projet 11-39553.93 et bourse 8210-043017).

Références

- Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5:183–207.
- Banel, M.H. & Bacri, N. (1997). Rôle des indices métriques et des indices phonotactiques lors de la segmentation lexicale en français. *L'Année Psychologique*, 97:77–112.
- Bradley, D.C. & Forster, K.I. (1987). A reader's view of listening. *Cognition*, 25:103–134.
- Burnage, G. (1990). *CELEX – A guide for users*, Rapp. tech., Centre for Lexical Information, University of Nijmegen, Nijmegen.

- Butterworth, B. (1983). Lexical representation, in B. Butterworth, éd., *Language Production, II: Development, Writing and Other Language Processes*, pp. 257–294, Academic Press, London.
- Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28:297–332.
- Church, K. (1987). Phonological parsing and lexical retrieval. *Cognition*, 25:53–69.
- Connine, C.M., Blasko, D.G., & Titone, D. (1993a). Do the beginnings of spoken words have a special status in auditory word recognition. *Journal of Memory and Language*, 32:193–210.
- Connine, C.M. & Titone, D. (1996). Phoneme monitoring. *Language and Cognitive Processes*, 11:647–654.
- Connine, C.M., Titone, D., & Wang, J. (1993b). Auditory word recognition: extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning Memory and Cognition*, 1:81–94.
- Content, A. & Frauenfelder, U.H. (1996). On the need for computer modeling: The case of language processing. *Psychologica Belgica*, 36:??
- Content, A., Mousty, P., & Radeau, M. (1990). Brulex: Une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique*, 90:551–556.
- Cutler, A. (1981). Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition*, 10:65–70.
- Cutler, A. & Butterfield, S. (1992). Rhythmic cues to speech segmentation — evidence from juncture misperception. *Journal of Memory and Language*, 31:218–236.
- Cutler, A. & Carter, D.M. (1987). The predominance of strong initial syllables in english vocabulary. *Computer Speech and Language*, 2:133–142.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language-specific comprehension strategy. *Nature*, 304:159–160.

- Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for syllable access. *Journal of Experimental Psychology: Human Perception and Performance*, 14:113–121.
- Dijkstra, T. & de Smedt, K., éd. (1996). *Computational Psycholinguistics: AI and Connectionist Models of Human Language Processing*, Taylor & Francis, London.
- Dupoux, E. & Mehler, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical codes. *Journal of Memory and Language*, 29:316–335.
- 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.
- Forster, K.I. (1976). Accessing the mental lexicon, in R.J. Wales & E.C.T. Walker, éd., *New Approaches to Language Mechanisms*, pp. 257–287, North-Holland, Amsterdam.
- Forster, K.I. (1979). Levels of processing and the structure of the language processor, in W.E. Cooper & E.C.T. Walker, éd., *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*, pp. 27–86, Lawrence Erlbaum, Hillsdale, New Jersey.
- Fowler, C.A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, 36:359–368.
- Fox, R.A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 10:526–540.
- Frauenfelder, U.H. (1991). Une introduction à la reconnaissance des mots parlés, in R. Kolinsky, J. Morais, & J. Segui, éd., *La reconnaissance des mots dans différentes modalités sensorielles. Données et modèles en psycholinguistique cognitive*, pp. 7–36, PUF, Paris.
- Frauenfelder, U.H. (1992). The interface between acoustic-phonetic and lexical processing, in M.E.H. Schouten, éd., *The Auditory Processing of Speech: From Sounds to Words*, Mouton de Gruyter, Berlin.

- Frauenfelder, U.H., Content, A., & Scholten, M. (in preparation). Lexical activation and deactivation in spoken word recognition. .
- Frauenfelder, U.H. & Kearns, R.K. (1996). Sequence monitoring. *Language and Cognitive Processes*, 11:665–673.
- Frazier, L. (1987). Structure in auditory word recognition. *Cognition*, 25:157–187.
- Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6:110–125.
- Gaskell, M.G., Hare, M., & Marslen-Wilson, W.D. (1995). A connectionist model of phonological representation in speech perception. *Cognitive Science*, 19:407–439.
- Goldman, J.-P., Content, A., & Frauenfelder, U.H. (1996). Comparaison des structures syllabiques en français et en anglais, in *XXIèmes Journées d'Étude sur la Parole*, pp. 119–122, Avignon, France.
- Grosjean, F. (1996). Gating. *Language and Cognitive Processes*, 11:597–604.
- Grosjean, F. & Frauenfelder, U.H. (1997). *A Guide to Spoken Word Recognition*, Psychological Press, London.
- Hardcastle, W.J. & Hewlett, N., éd. (in press). *Instrumental studies of coarticulation*, Cambridge University Press, Cambridge, UK.
- Harnad, S., éd. (1987). *Categorical Perception: The Groundwork of Cognition*, Cambridge University Press, Cambridge, UK.
- Kolinsky, R. (1998). Spoken word recognition: A stage-processing approach to language differences. *European Journal of Cognitive Psychology*, 10:1–40.
- Kuhl, P. (1991). Human adults and human infants show a ‘perceptual magnet effect’ for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50:93–107.

- Kutas, M. & van Petten, C.K. (1994). Psycholinguistics electrified: Event-related brain potential investigations, in M.A. Gernsbacher, éd., *Handbook of Psycholinguistics*, pp. 83–113, Academic Press, San Diego.
- Lahiri, A. & Marslen-Wilson, W. (1991). The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition*, 38:245–294.
- Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica*, 5:1–54.
- Lieberman, A.M. (1996). *Speech: A Special Code*, MIT Press, Cambridge, Mass.
- Luce, P.A., Pisoni, D.B., & Goldinger, S.D. (1990). Similarity neighborhoods of spoken words, in G.T.M. Altmann, éd., *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, pp. 122–147, MIT Press, Cambridge.
- 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. (1984). Function and process in spoken word recognition, in H. Bouma & D.G. Bouwhuis, éd., *Attention and Performance X: Control of Language Processes*, p.??, Lawrence Erlbaum Associates, Hillsdale, NJ.
- Marslen-Wilson, W.D. & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8:1–71.
- Marslen-Wilson, W.D. & Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology*, 10:29–63.
- Marslen-Wilson, W.D. & Zwitserlood, P. (1989). Accessing spoken words: the importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15:576–585.
- Massaro, D.W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*, MIT Press, Cambridge, Mass.

- Massaro, D.W. & Cohen, M.M. (1983). Categorical or continuous speech perception: a new test. *Speech Communication*, 2:15–35.
- Mattys, S.L. (1997). The use of time during lexical processing and segmentation: a review. *Psychonomic Bulletin & Review*, 4:310–329.
- McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18:1–86.
- McQueen, J.M. (1996). Word spotting. *Language and Cognitive Processes*, 11:695–699.
- McQueen, J.M. & Cox, E. (1995). The use of phonotactic constraints in the segmentation of dutch, in *Proceedings of Eurospeech '95*, vol. 3, pp. 1707–1710, Madrid, Spain.
- McQueen, J.M. & Cutler, A. (1997). Morphology in word recognition, in A.M. Zwicky & A. Spencer, éd., *The Handbook of Morphology*, p.??, Blackwell, Oxford.
- McQueen, J.M., Norris, D., & Cutler, A. (1994). Competition in word recognition — spotting words in other words. *Journal of Experimental Psychology: Learning Memory and Cognition*, 20:621–638.
- Mehler, J. (1981). The role of syllables in speech processing: Infant and adult data. *Philosophical Transactions of the Royal Society, Series B*, 295:333–352.
- Mehler, J., Dommergues, J., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20:298–305.
- Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7:323–331.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76:165–178.
- Nakatani, L.H. & Dukes, K.D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, 62:714–719.

- Norris, D. (1992). Connectionism: A new breed of bottom-up model, in R. Reilly & N. Sharkey, éd., *Connectionist approaches to natural language processing*, Lawrence Erlbaum, Hove, UK.
- Norris, D., McQueen, J.M., & Cutler, A. (????). Merging phonetic and lexical information in phonetic decision-making, submitted for publication.
- Norris, D., McQueen, J.M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning Memory and Cognition*, 21:1209–1228.
- Norris, D. J. & Cutler, A. (1988). The relative accessibility of phonemes and syllables. *Perception & Psychophysics*, 43:541–550.
- Palmeri, T.J., Goldinger, S.D., & Pisoni, D.B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning Memory and Cognition*, 19:309–328.
- Perkell, J.S. & Klatt, D.H., éd. (1986). *Invariance and Variability in Speech Processes*, Lawrence Erlbaum, Hillsdale, N.J.
- Peterson, G.E. & Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24:175–184.
- Pisoni, D.B. (1993). Long-term memory in speech perception — some new findings on talker variability, speaking rate and perceptual-learning. *Speech Communication*, 13:109–125.
- Pisoni, D.B. & Luce, P.A. (1987). Acoustic-phonetic representations in word recognition, in U.H. Frauenfelder & L.K. Tyler, éd., *Spoken word recognition*, pp. 21–52, MIT Press, Cambridge, Mass.
- 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.

- Samuel, A. (1981). Phoneme restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, 110:474–494.
- Savin, H.B. (1963). Word frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America*, 35:200–206.
- Stevens, K.N. (1986). Models of phonetic recognition II: A feature-based model of speech recognition, in *Proceedings of the Montreal Satellite Symposium on Speech Recognition, XIIth International Congress on Acoustics*, pp. 66–67.
- Suomi, K., McQueen, J.M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, 36:422–444.
- Swinney, D.A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18:645–660.
- Tabossi, P. (1996). Cross-modal semantic priming. *Language and Cognitive Processes*, 11:569–576.
- Taft, M. & Forster, K.I. (1976). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14:630–647.
- Tanenhaus, M.K., Leiman, J.M., & Seidenberg, M.S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Memory and Verbal Behavior*, 18:427–440.
- Tyler, L.K., Marslen-Wilson, W.D., J., Rentoul, & Hanney, P. (1988). Continuous and discontinuous access in spoken word recognition: The role of derivational affixes. *Journal of Memory and Language*, 27:368–381.
- Warren, P. & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, 41:262–275.
- Warren, P. & Marslen-Wilson, W. (1988). Cues to lexical choice - discriminating place and voice. *Perception & Psychophysics*, 43:21–30.

Warren, R. (1970). Perceptual restoration of missing speech sounds. *Science*, 167:392–393.

Zwitserslood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word recognition. *Cognition*, 32:25–64.

Figure 1: Représentations et processus de traitement dans la reconnaissance des mots (perspective classique).

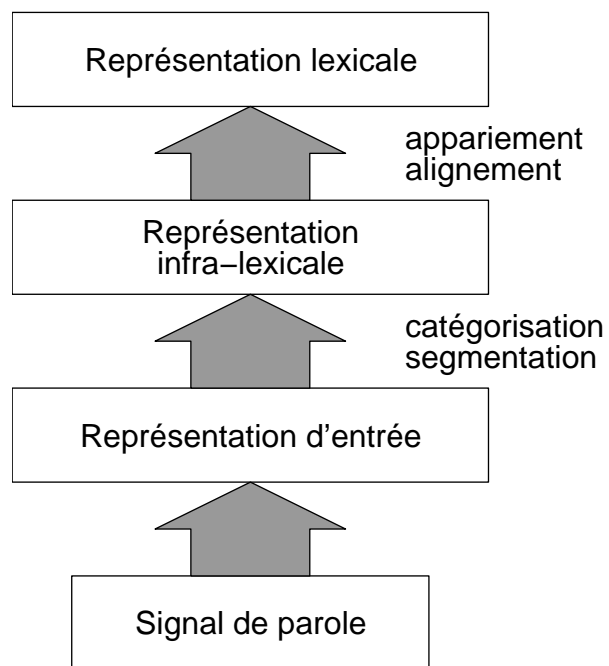
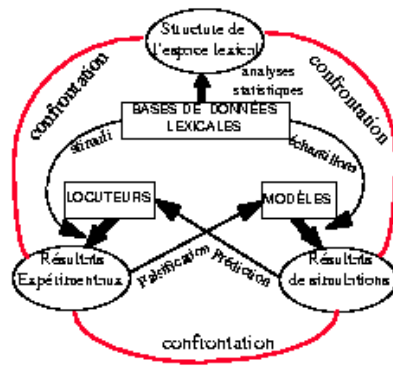


Figure 2: Interactions entre les différentes méthodes employées dans les études sur la reconnaissance des mots.



Légendes des figures

1. Représentations et processus de traitement dans la reconnaissance des mots (perspective classique).
2. Interactions entre les différentes méthodes employées dans les études sur la reconnaissance des mots.