

**Sophie WAUQUIER-GRAVELINES**  
**Université de Nantes**

**SEGMENTATION LEXICALE DE LA PAROLE CONTINUE  
LA LINÉARITÉ EN QUESTION**

**RÉSUMÉ**

Les caractéristiques physiques du signal, ainsi que les modélisations de la chaîne parlée le plus souvent retenues dans les modèles psycholinguistiques (représentation orthographique ou phonologique SPE) ont longtemps influencé une conception strictement algorithmique et linéaire du traitement de la parole continue. Néanmoins, la synthèse des plus récentes propositions de la psycholinguistique expérimentale et de la phonologie multilinéaire nous amène à mettre à jour les limites d'une telle conception et nous invite à envisager l'existence de stratégies heuristiques incluant le stockage temporaire et/ou l'existence de calculs rétroactifs permettant l'extraction rapide et efficace de l'information lexicale.

**MOTS-CLÉS**

Segmentation, linéarité, traitement de la parole continue, phonologie.

De la linéarité du signifiant, établie en principe par Ferdinand de Saussure et systématisée par les phonologues structuralistes, aux débats les plus actuels concernant la nécessité de l'existence de positions squelettales dans les représentations multilinéaires, la phonologie se voit confrontée à intervalles réguliers au problème de la continuité et de la linéarité du signal de parole. Elle voit s'opposer à des formalisations abstraites l'inertie entêtante de la réalité physique dont elle a à rendre compte et qui l'amène à reposer régulièrement la question du rapport entre forme et substance. Les modèles psycholinguistiques rendant compte du traitement de la parole continue, et en particulier les modèles d'accès au lexique, sont de ce point de vue intéressants. Dans la mesure où l'accès lexical modélise le passage d'une information concrète acoustiquement présente dans le signal à des représentations abstraites, il pose la question du rapport entre substance et forme d'une manière qui n'est pas la plus usitée. En effet, la linguistique théorique a pour tradition de proposer quasi uniquement des modèles de production, explicitant le passage de la forme à la substance et abordant la question des représentations cognitives essentiellement d'un point de vue statique. Dans la plupart des modèles phonologiques, les connaissances sont représentées comme un savoir structuré statique ou configurationnelles (Laks, 1996), stocké dans le lexique et donnant lieu à des aménagements (application de règles, de contraintes, rattachement de positions, etc.). Cette compétence phonologique, ainsi formalisée intègre néanmoins une dimension temporelle qui induit une représentation linéaire (matrices de traits, existence d'un squelette). Les modèles d'accès lexical au contraire rendent compte du traitement à partir du signal et posent le problème de la représentation des connaissances en termes de savoirs procéduraux (traitement, décodage, « *parsing* »). La question de la linéarité est alors abordée par le biais des problèmes que pose la temporalité physique de la substance sonore. Linéarité du signe, ou linéarité du signal, phonologie et accès au lexique posent par conséquent le problème de deux points de vue différents. Les modèles d'accès lexical, en nous invitant à réfléchir sur le traitement perceptif de la linéarité du signal de parole, nous permettent de nous interroger sur le passage de la substance à la forme, tâches incombant également au linguiste, me semble-t-il.

## **1. Segmentation lexicale et traitement de la parole continue**

### **1.1. L'accès au lexique : reconnaître les mots dans la chaîne**

La question du traitement de l'entrée parlée, quand elle rend compte des processus de traitement se situant au-delà du décodage des informations acoustiques et de la normalisation phonétique (débit, timbre, durée, etc.), a donné lieu à de nombreuses recherches concernant l'accès au lexique. L'accès au lexique est considéré comme le premier stade d'interface entre information acoustico-phonétique et traitement linguistique permettant à un auditeur d'interpréter sémantiquement et syntaxiquement la chaîne parlée. Il désigne les processus

mentaux complexes qu'utilise un auditeur pour atteindre et mettre en œuvre les informations lexicales dont il dispose, lors de la compréhension d'un énoncé oral. Les processus mentaux ainsi envisagés reposent sur l'hypothèse de l'existence d'un «lexique mental» comportant toutes les unités lexicales connues. Ces unités auraient une représentation multidimensionnelle incluant des informations orthographiques, phonologiques, morphologiques, syntaxiques et sémantiques. L'accès lui-même est considéré comme un appariement entre les représentations stockées dans le lexique interne et les unités dégagées de l'analyse sensorielle du signal. Il est généralement modélisé en trois étapes (Frauenfelder & Tyler, 1987) : 1) la recherche de l'information pertinente dans le signal, 2) l'activation des mots-candidats possibles correspondant à l'information contenue dans le signal, 3) la sélection définitive du mot recherché ainsi que la résolution des ambiguïtés éventuelles de découpage du signal. Les modélisations respectives de ces trois étapes sont bien évidemment dépendantes les unes des autres, l'activation et la procédure de sélection du mot-candidat étant consécutives au traitement initial du signal de parole. Les modèles, ainsi que les propositions théoriques qu'ils sous-tendent et qui les contraignent, vont diverger quant à la manière dont ils vont définir ce qui constitue l'information pertinente contenue dans le signal. Autrement dit, quant à la manière dont ils vont répondre à la question de la temporalité et de la linéarité du signal.

## **1.2. Des stratégies de segmentation : traitement séquentiel ou recherche des informations saillantes ?**

Pour des raisons anatomiques, la parole, succession de gestes articulatoires, est produite séquentiellement et contrainte par un facteur temporel. Les sons de parole sont produits successivement et disponibles linéairement pour l'auditeur. Si l'auditeur entend une suite comme [ɛ̃gRãanimal], l'information acoustique disponible lui permettant d'accéder à une unité lexicale possible donc correspondant à [ɛ̃], précède nécessairement [g], [ã] ou [t]... De même, il entendra la syllabe [ɛ̃] avant la syllabe [gRã] et le mot « grand » avant le mot « animal ». Pour les mêmes raisons, l'information acoustique n'est physiquement disponible que pendant une durée très courte et ne peut plus être récupérée en cas de perception erronée. Recevant cette information acoustique continue et linéaire, comment l'auditeur procède-t-il pour y récupérer des unités significatives lui permettant de comprendre le message oral ?

Les modèles d'accès au lexique vont apporter diverses réponses à ce problème complexe. On peut considérer qu'ils se répartissent sur un continuum dont l'un des extrêmes propose une conception strictement linéaire du traitement. Pour cette première famille de modèles, l'hypothèse retenue envisage que la succession d'événements articulatoires est lisible en une succession d'événements acoustiques disponibles dans le signal et que le flux d'informations délivré séquentiellement induit automatiquement chez l'auditeur un décodage linéaire de

l'information. En d'autres termes l'information est décodée en temps réel de manière séquentielle. Dans un énoncé tel que [ɛgRātanimal] « un » sera nécessairement traité avant « grand » lui-même traité avant « animal ». Selon les modèles, les divergences porteront sur la nature et la taille de l'unité d'accès : phonème, syllabe ou mot.

L'autre manière d'envisager la linéarité du signal consiste à considérer la chaîne parlée non pas comme une suite d'unités isomorphes mais comme une succession de contrastes faisant alterner des endroits porteurs d'informations saillantes (informations prosodiques telles que l'accent par exemple) et des endroits qui, ne comportant aucune information saillante, peuvent, par conséquent, être négligés ou traités secondairement. Cette conception suppose alors que l'accès lexical n'est pas opéré linéairement, mais que la stratégie de segmentation donnera priorité aux portions de signal porteuses d'informations saillantes même si elles sont produites après des portions de signal ne comportant pas ce type d'informations. Dans un énoncé tel que [ɛgRātanimal] « animal » est doublement saillant. Syntaxiquement, il est tête du syntagme nominal. Prosodiquement, il est porteur d'informations sur la dernière syllabe qui indiquent la limite droite du GN. Dans une phrase assertive par exemple, cette syllabe sera significativement plus longue, affectée d'une baisse d'intensité ainsi que d'une chute de la F0. Dans cette deuxième série de modèles, l'hypothèse retenue serait la suivante : la syllabe finale du mot de contenu aura un rôle déterminant dans l'accès lexical qui s'effectuera de manière régressive à partir d'elle. L'accès au mot de contenu « animal » sera réalisé avant l'accès au déterminant, mot grammatical clitique, et avant l'accès à l'adjectif, même si ils sont acoustiquement disponibles avant la syllabe saillante.

Ces deux conceptions opposées de l'accès au lexique, outre les différences concernant proprement la modélisation de processus psychologiques mettant en cause des divergences théoriques structurant le champ de la psycholinguistique <sup>1</sup>, reposent deux questions tout à fait centrales en linguistique.

En premier lieu, ils posent différemment la question du rapport entre forme et substance. Dans la première famille de modèles, la différence entre linéarité de la forme et linéarité de la substance n'est pas clairement affirmée. La définition des unités ou des domaines de l'accès n'explique pas toujours le niveau d'abstraction des représentations concernées et la plupart des modèles invitent implicitement à considérer le signal de parole comme une suite d'unités organisées (suite de phonèmes, suite de syllabes, suite de mots). Dans la deuxième famille de modèles, au contraire, on postule que le traitement de la substance envisagée de manière plus précise et plus réaliste, puisqu'il intègre le décodage des informations suprasegmentales telles que intensité, débit, accents, allongements, baisse de la F0, permet ensuite d'accéder à des unités structurelles organisées. Ceci suppose alors qu'il existe une différence de nature entre substance et forme, qui doivent être mises en relation par une ou plusieurs opérations mentales qui restent à définir. Les questions encore largement non résolues que cette conception soulève sont les suivantes :

Quelles sont alors les procédures d'accès ? Jusqu'à quel point peut-on les considérer comme heuristiques ?

Doit-on envisager ces procédures comme des opérations inverses de celles que propose la phonologie – règles, applications successives de contraintes – et considérer par conséquent que les modèles de perception et les modèles de production sont identiques et réversibles ?

Si ce n'est pas le cas comment pouvons-nous alors envisager le rapport particulier qu'entretiennent phénomènes perceptifs et représentations phonologiques abstraites ?

La deuxième question que cette opposition met en lumière est celle de l'universalité des procédures d'accès. Dans le premier cas, si l'accès au lexique se fait de manière strictement séquentiel, toutes les langues vont donner lieu à des procédures d'accès identiques. Seule variera éventuellement l'unité d'accès (certaines langues traiteront phonème par phonème ou d'autres syllabe par syllabe), mais la dynamique linéaire restera la même. Dans le deuxième cas au contraire, dans la mesure où la nature de l'information saillante a de fortes chances de varier d'une langue à l'autre, on doit alors considérer que les procédures d'accès vont être différentes et que les paramètres phonologiques propres à chaque langue vont contraindre les opérations mentales permettant le passage de la substance à la forme.

Dans les deux sections suivantes nous présentons ces deux familles de modèles, les réponses apportées à la question de la linéarité et les limites qu'elles comportent, tant du point de vue psycholinguistique que du point de vue phonologique.

## **2. Modèles linéaires et modèles postlexicaux**

Comme nous l'avons vu, les modèles linéaires supposent que la segmentation est strictement séquentielle et orientée de gauche à droite (Cole et Jakimik, 1978, 1980 ; Marslen-Wilson et Welsh, 1978 ; Nootboom, 1981) et que la frontière gauche d'un mot est déductible de la reconnaissance du mot qui le précède<sup>2</sup>. Ils postulent donc d'une part, que les mots sont reconnus séquentiellement dans l'ordre où ils se présentent quelles que soient les relations syntaxiques et sémantiques qui les unissent, d'autre part que les auditeurs identifient un mot avant qu'il ne soit terminé : ceux-ci doivent effectivement reconnaître le mot qui défile assez rapidement pour pouvoir se préparer à traiter celui qui suit. Parmi les modèles de ce type, le plus représentatif est le modèle « COHORT » (Marslen-Wilson et Welsh, 1978 ; Marslen-Wilson, 1987) qui a fortement influencé la conception linéaire de l'accès au lexique sur la base de très nombreuses validations expérimentales obtenues avec des tâches variées. Ce modèle « COHORT » de (Marslen-Wilson et Welsh, 1978) comporte deux étapes distinctes. Lors d'une première étape, le système active une « cohorte » contenant tous les mots du lexique correspondant aux 200 premières millisecondes du signal (environ à la première syllabe d'un mot). Puis, au fur et à mesure que

l'information acoustico-phonétique est délivrée, les mots qui ne correspondent plus à cette information disparaissent de la cohorte par désactivation. Le mot est reconnu quand son « point d'unicité » (PU) est atteint, quand il reste le seul dans la cohorte. Ainsi, si le mot à reconnaître est par exemple « éléphant », les mots « élaborer », « élargir », « élastique », « élection », etc. vont être activés. Puis, tous les mots incompatibles avec l'information qui est peu à peu délivrée sont désactivés ; « éléphant » est reconnu quand les informations acoustiques correspondant à la fricative labiodentale « f » apparaît, il reste seul dans la cohorte.

Le modèle impose un ajustement strict entre le signal et la représentation lexicale : la possibilité de reconnaître un mot est strictement soumise à l'encodage juste de l'information correspondant au début de ce mot. Si la cohorte est constituée sur la base de la représentation acoustico-phonétique des unités lexicales, elle est contrainte par les variations phonétiques et phonologiques se produisant dans le signal. Pour résoudre ce problème, et limiter l'importance drastique du début de mot, dans une deuxième version du modèle Cohort II, Warren et Marslen-Wilson (1987, 1988), Lahiri et Marslen-Wilson (1991) proposent que la cohorte ne soit pas directement dérivée de l'information acoustique mais de la représentation phonologique des mots en termes de traits : l'auditeur utilise l'information acoustique pour en dériver une cohorte de représentations sous-spécifiées. Ceci permet d'inclure dans la cohorte des variantes libres ou complémentaires du ou des phonèmes initiaux concernés et donc d'élargir la cohorte initiale à toutes les variantes possibles du mot.

Ce modèle ou cette famille de modèles présentent divers avantages. Ils proposent d'abord une procédure algorithmique simple et sont cognitivement peu coûteux. Une telle conception de l'accès n'impose en effet ni la formalisation d'une compétence phonologique, morphologique et syntaxique complexe ni la nécessité de stockage d'informations, de retours en arrière, et de calculs utilisant cette compétence. L'argument de la simplicité n'est pas trivial. La vitesse à laquelle se réalise la compréhension en temps réel d'un énoncé oral amène à considérer que le traitement se fait de manière relativement automatisée et directe. D'autre part, cet algorithme est applicable à n'importe quelle langue sans qu'il soit nécessaire d'y apporter des ajustements importants. Il propose donc une procédure universelle conforme aux exigences de généralisation imposées par une démarche scientifique.

L'hypothèse de traitement séquentiel proposé par le modèle COHORT a été largement testée par l'obtention d'effets montrant l'existence effective et le rôle du « PU ». En effet si l'accès au lexique se produit bien de manière séquentielle et linéaire on peut penser que les mots ayant un « PU » précocement placé dans le mot (sur la première voyelle ou la première syllabe) seront reconnus avant des mots ayant un « PU » tardif (dernière syllabe par exemple). Des effets de ce type ont été massivement observés avec des tâches variées : détection de phonèmes (Frauenfelder, Segui et Dijkstra, 1990 ; Marslen-Wilson, 1984), décision lexicale (Taft et Hambly, 1986), *gating* (Grosjean, 1980), etc. (Pour une synthèse voir Wauquier-Gravelines 1996 ; Mattys, 1997.) Les résultats suggèrent donc que les

effets de « PU » sont robustes et que l'hypothèse d'une stratégie linéaire d'accès au lexique peut être sérieusement envisagée.

Néanmoins, Radeau Mousty et Padeloup (1994) ont remis en cause ces effets en montrant qu'en accélérant le débit du matériel acoustique proposé jusqu'à le rapprocher de la valeur du débit de la parole spontanée, on fait disparaître les effets de « PU ». On peut donc envisager qu'il y ait un traitement séquentiel à débit lent mais que cette procédure algorithmique simple ne suffit pas à rendre compte du traitement de la parole naturelle. Pour un débit de parole spontanée, le traitement séquentiel doit être réenvisagé.

D'autres arguments expérimentaux ont en outre remis en cause cette conception purement séquentielle. Une analyse statistique du lexique a montré que le point d'identification de la plupart des mots courts, qui sont souvent les plus fréquents, se trouve après leur dernier segment. Seuls 40 % des mots du lexique pourraient être reconnus sans ambiguïté avant la fin (Luce, 1986). Grosjean<sup>3</sup> (1985), Bard, Shillcock et Altmann (1989) ont également montré, en utilisant la technique de gating, qu'un nombre non négligeable de mots, en particulier les mots grammaticaux et les mots courts, sont reconnus après la fin de leur réalisation acoustique. Leurs résultats indiquent d'une part que la procédure de recherche d'une unité lexicale n'impose pas nécessairement que l'unité précédente ait été reconnue, et d'autre part que les auditeurs peuvent faire simultanément plusieurs hypothèses de segmentation en parallèle qui sont définitivement désambiguïsées par des informations phonologiques, morphologiques, syntaxiques voire pragmatiques. L'existence de procédures d'accès concurrentes qui pourraient être calculées parallèlement soulève une autre question posée par la linéarité et à laquelle les modèles purement linéaires ne répondent pas. Le signal peut comporter des chaînes ambiguës présentant plusieurs possibilités de segmentation. Les cas les plus courants concernent les mots enchâssés (une suite comme [kR:ʁodil] peut être segmentée en « crocodile » ou « croque Odile » ou « croque », « ode », « île » ; « élémentaire » contient « élément » dont il est le dérivé) mais elles peuvent aussi se produire entre les mots à l'intérieur d'une phrase ou d'un syntagme. Une suite comme [pøtitami] peut être segmentée « petite amie », « petit ami », « petit tamis ».

Une autre génération de modèles d'inspiration connexionniste, dits interactifs, *Trace* (Mc Clelland et Elman, 1986), *Shorlist* (Norris, 1994) et *Cohort II* apportent une réponse à ce type de problèmes. Ces trois modèles ont en effet en commun de ne pas supposer un alignement strict entre l'information sensorielle et les unités lexicales et de permettre au système de maintenir simultanément plusieurs hypothèses de segmentation dont la résolution est obtenue par une dynamique d'activation et d'inhibition simultanée entre divers types d'informations distribuées dans le système. Dans un modèle comme *Trace*, par exemple, chaque unité à chaque niveau est représentée par un nœud qui se caractérise 1) par un seuil d'activation au-delà duquel le nœud est dit actif et 2) un niveau de repos en l'absence d'entrées extérieures. Chaque nœud est relié aux autres nœuds d'un même niveau par des connexions inhibitrices, et aux nœuds des

unités des autres niveaux par des connexions facilitatrices. La reconnaissance d'un mot procède de la dynamique d'activation (activation multiple) et d'inhibition (inhibition latérale) entre les différentes unités, consécutive à l'apparition d'une entrée extérieure. Dans ce type de modèles, le système ne démarre pas son processus de recherche à partir d'un point spécifique. Le nombre des candidats lexicaux n'est donc pas restreint par la correspondance stricte à un moment donné entre l'entrée sensorielle et les unités lexicales ou infralexicales. Tous les mots du lexique, mais aussi toutes les séquences sonores possibles dans la langue, correspondant à l'entrée sensorielle sont susceptibles d'être activés. Ainsi, par exemple, si le mot présenté est « crocodile », le système va activer tous les mots commençant par « cro » (« croc », « crocodile », « crochet », « crotale ») mais également, « crocodilien », « eau », « ode », « Odile », « île » qui vont être très rapidement inhibés parce qu'ils ne correspondent pas longtemps à l'entrée sensorielle. La segmentation est donc consécutive à la reconnaissance des mots mais sans qu'aucun alignement positionnel ne soit nécessaire. Chaque séquence à chaque instant est envisagée comme une unité lexicale. Seul le seuil d'activation d'une unité lexicale la distingue de toutes les autres. Son émergence impose consécutivement les frontières qui le séparent du mot précédent et du mot suivant. La segmentation est alors résiduelle.

Ces modèles de deuxième type tout en dépassant les contradictions des modèles strictement linéaires conservent les qualités de simplicité et rapidité de ceux-ci. Néanmoins, ils ne remettent pas fondamentalement en cause la linéarité des procédures d'accès. Le traitement se fait toujours strictement de gauche à droite, l'originalité consistant essentiellement à maintenir en parallèle et simultanément toutes les possibilités de découpage linéaire et à introduire la possibilité d'un délai temporel entre l'arrivée de l'information et la sélection lexicale. La critique majeure qui a été faite à un modèle tel que « Trace » est la suivante : l'alignement exhaustif supposant toutes les segmentations possibles à chaque instant crée un modèle surgénératif. Le nombre d'hypothèses, et notamment d'hypothèses fausses, qu'il doit envisager pour trouver chaque mot a occasionné des critiques vives (Cutler et Butterfield, 1992 ; Christophe, 1993) ; et suppose que le système possède une puissance de calcul, qui sert en grande partie à éliminer des hypothèses inutiles <sup>4</sup>. Des recherches expérimentales en amorçage intermodal <sup>5</sup> (Gow et Gordon, 1995) ont validé cette hypothèse d'une activation parallèle des segmentations lexicales possibles dans le signal. On émettra quelques réserves quant à la pertinence de la tâche : l'amorçage intermodal ne teste pas directement les hypothèses de segmentation, mais on peut considérer qu'un modèle à activation parallèle de type *Trace* permet effectivement de rendre compte de la segmentation de la parole continue et de répondre le plus souvent aux problèmes posés par la linéarité.

Néanmoins, si les solutions apportées par cette deuxième génération de modèles linéaires semblent satisfaisantes d'un point de vue psycholinguistique, on peut leur opposer d'un point de vue linguistique les mêmes critiques qu'aux modèles purement séquentiels.



En premier lieu, dans tous les cas, la conception linéaire de l'accès semble résulter non pas d'une observation directe du signal mais d'une projection dans le signal de la linéarité des représentations abstraites, faisant correspondre strictement linéarité du signe et linéarité du signal, autrement dit consistant à considérer que la linéarité du signal résulte en quelque sorte de la linéarité du signe. Or, ceci revient, en quelque sorte à poser le problème à l'envers. Ces modèles séquentiels, en proposant un alignement de l'information sensorielle contenue dans le signal et des unités abstraites de la grammaire mentale qu'utilise l'auditeur, supposent plus ou moins implicitement une isomorphie entre substance et forme. Or, si le signal peut se lire comme une succession d'événements acoustiques, on ne trouve pas de corrélats acoustiques d'unités abstraites infralexicales décrites par la phonologie. Les phones ne sont pas produits de manière isolée et les analyses spectrales de la chaîne sonore montrent clairement que la plupart des phones sont influencés par les autres phones adjacents quelle que soit leur position dans la chaîne parlée (Liberman, 1970 ; Studdert-Kennedy, 1975). La chaîne ne peut donc pas être envisagée *a priori* ni comme une suite d'objets abstraits (phonèmes), ni comme une suite d'événements acoustiques correspondant à des objets abstraits. Or on retrouve dans tous ces modèles un niveau segmental dont l'activation des unités aboutit à la reconnaissance du mot. Dans tous les cas, l'auditeur procède de gauche à droite et traite linéairement des tranches de signal assimilés à des « phonèmes », même si comme dans *Trace* ou *Cohort II*, il maintient simultanément des hypothèses concurrentes.

Malgré la définition du lexique mental donnée plus haut, considérant que « les représentations lexicales sont nécessairement multidimensionnelles » (Lecocq et Segui, 1989), ces modèles projettent sur le signal une représentation de la compétence phonologique en définitive extrêmement primitive. Cette compétence phonologique qui serait mise en œuvre par les auditeurs lors de l'accès au lexique consiste en une transcription phonétique des mots orthographiques, avec un seul niveau de représentation comportant des unités phonématiques. Ces représentations sont unilinéaires et plates, et se résument, en fait, à une compétence orthographique linéaire et segmentale adaptée à la modalité auditive. Elles ne correspondent à aucune réalité physique. Ces représentations, comme l'ont souligné Grosjean et Gee (1987) donne une place centrale au mot alphabétique et le considère nécessairement soit comme *unité d'accès au lexique* ou soit comme *domaine de l'accès* adoptant en cela une attitude conservatrice. Cette représentation de la compétence phonologique est aujourd'hui théoriquement dépassée.

En second lieu, cette conception omet totalement l'existence des informations suprasegmentales. Or la présence de l'information suprasegmentale permet que la production orale soit perçue comme une structure claire, bien intonée, rythmée, dynamique. Elle participe pleinement à l'intelligibilité du message contenu par le signal. En adoptant une conception alphabétique des unités de traitement de la parole continue, le cadre conceptuel proposé pour rendre compte de l'accès lexical en parole continue privilégie une notation et une

conception de la parole qui excluent ces informations. Ce fait pose un double problème :

D'une part, aucune justification théorique n'est avancée pour défendre une conception de la parole qui néglige toute une partie de l'information qui lui est pourtant spécifique. Les difficultés et les limites qu'impose l'absence de représentation de la prosodie sont très rarement évoquées. On se trouve alors face à une formulation tautologique des hypothèses de traitement de la parole continue : on ne peut envisager l'accès lexical en parole continue autrement que comme l'extrapolation de résultats concernant la reconnaissance des mots isolés.

Ce choix pose d'autre part un problème expérimental. Cette information est négligée théoriquement, alors qu'elle est présente dans le matériel testé, en particulier quand les tâches de détection ou de reconnaissance lexicale sont réalisées en phrases. Ce fait constitue un risque de biais majeur dans l'interprétation des données obtenues : la part que l'information prosodique peut avoir dans les performances des sujets constitue rarement un facteur expérimental, alors que ceux-ci reçoivent cette information et peuvent s'en servir.

Enfin, cette conception de la linéarité du signe posée par Saussure, puis amplifiée et systématisée par la linguistique structurale puis par la phonologie modèle SPE, est directement inspirée d'un modèle alphabétique et « eurocentrique » (Cao Xan Hao, 1985) qui suppose d'une part que les notions de phonèmes, de mots sont claires et d'autre part, que le mot ainsi défini existe dans toutes les langues. Or, si cette unité est utilisée dans la plupart des descriptions linguistiques comme l'unité qui sépare les constructions infra-lexicales et la syntaxe, il n'en existe aucune définition satisfaisante et généralement acceptée. Notre propos ne concernant pas la définition linguistique de l'unité lexicale, nous ne multiplierons pas les exemples et ne détaillerons pas les diverses définitions (morphologiques, sémantiques, syntaxiques, etc.) du mot proposées par la théorie linguistique (Matthews, 1974 ; Spencer, 1991). Ce qui nous importe ici est de souligner que la conception alphabétique du mot est, selon nous, la moins susceptible de correspondre à une unité de traitement universellement utilisée.

Enfin, si cette représentation phonologiquement improbable permet dans un modèle à activation multiple de résoudre le problème de la segmentation des mots enchâssés, elle ne permet pourtant ni aux modèles strictement linéaires, ni aux modèles à activation multiple de rendre compte du traitement en temps réel de la parole continue dès lors que s'y produisent des phénomènes phonologiques tels que la liaison (Wauquier-Gravelines, 1996).

Un modèle comme *Trace* ne peut pas rendre compte aisément du traitement de la liaison enchaînée. La suite /pətitami/ donne lieu à l'activation de « petite », « petit » « amie » « ami » et « tamis ». Mais dans la mesure où *Trace* est un modèle qui suppose une représentation linéaire et « alphabétique » des unités lexicales, la représentation sous-jacente de l'adjectif est la suivante : /pəti/ au masculin /pətit/ au féminin. Le « [t] » présent dans la chaîne va activer soit la forme féminine de l'adjectif /pətit/ soit le mot /tami/. Une suite comme [pətitami] ne peut donc être segmentée que de deux manières : « petite amie » correspondant

à /pətɪt/ + /ami/ ou « petit tamis » correspondant à /pətɪ/ + /tami/. La segmentation « un petit ami » avec liaison est impossible car les représentations sous-jacentes ne portent pas trace du « t » flottant (/pətɪ/ + /ami/). Si l'entrée est [pøtitami] la forme masculine sera inhibée systématiquement par la forme féminine, on obtiendra alors la segmentation « un petite amie » qui est incorrecte. Le problème n'est donc pas prioritairement de désambiguïser « petit ami » et « petit tamis », mais de récupérer « petit ami ». Aucune suite contenant une liaison enchaînée (« petit éléphant », « grand âne », etc.) ne peut être segmentée par *Trace* si l'on n'admet pas qu'il faut adjoindre au modèle une représentation sous-jacente /pəti + (t)/ permettant de savoir que l'adjectif au masculin peut se réaliser soit avec un /t/ soit sans /t/, c'est-à-dire une formalisation de la consonne flottante. Cette remarque vaut également pour un modèle tel que « Cohort I » où toutes les hypothèses de segmentation d'une suite comme « un petit éléphant » s'avèrent problématiques. La première possibilité suppose que l'auditeur segmente avant l'apparition du /t/ : quand il entend [pøti], il reconnaît « petit » et ouvre la cohorte suivante des mots commençant par [te] (téléphone, téléphérique, terrier). Il ne peut retrouver « éléphant » qu'après avoir constaté que cette hypothèse était fautive. L'autre possibilité serait de segmenter après le /t/, quand l'auditeur entend [pøtit], il reconnaît « petite » dont le genre n'est pas conforme à celui de l'article, ensuite il initialise une cohorte où il a déjà plus de chance de trouver « éléphant » (puisque la cohorte n'est pas contrainte par l'expression du genre et du nombre par l'article), mais il sélectionne alors une suite agrammaticale « un petite éléphant ».

Racine et Grosjean (à paraître) sont arrivés à des conclusions similaires concernant l'effacement du schwa sur frontières lexicales. On peut s'attendre à ce que tous les phénomènes phonologiques entraînant une modification ou une réorganisation syllabique de la frontière gauche des unités lexicales (élisions, mutations présentes dans certaines langues celtiques, etc.) mettent ces modèles en difficulté et reposent la question de la linéarité du traitement.

### 3. Modèles prélexicaux

La deuxième famille de modèles que nous avons évoquée repose à l'inverse sur une conception prélexicale de la segmentation s'appuyant sur des informations acoustiquement saillantes. Les hypothèses concernant une stratégie de segmentation prélexicale ont été, jusqu'à maintenant, l'objet de beaucoup moins de recherches que les modèles postlexicaux. Ceci peut être attribué d'abord au fait que l'hypothèse d'une segmentation prélexicale impose, en même temps que la modélisation d'une stratégie de traitement, la définition de l'information pertinente sur laquelle s'appuient les auditeurs pour opérer la segmentation. Cette stratégie de segmentation ne peut être efficace qu'à condition qu'il existe dans le signal des traces correspondant à des frontières lexicales ou des indices permettant à l'auditeur de développer des inférences sur la place éventuelle d'une frontière. De nombreuses recherches pour des langues assez variées ont tenté de mettre en évidence l'existence d'informations spécifiques délimitant des unités de

traitement permettant de faciliter l'accès au lexique. L'existence d'une part d'indices acoustiques tels que les variations de la F0 (Vaissière 1995a, 1995b), ou l'allongement consonantique et vocalique (Oller, 1973 ; Umeda, 1975, 1977 ; Lehiste, 1965, 1966 ; Quené, 1992 ; Christophe, 1993), d'autre part de phénomènes prosodiques tels que l'accent (Vaissière, 1995 ; Cutler & Carter, 1987) a été mise en évidence dans un très grand nombre de travaux. Les résultats de ces recherches sont pourtant extrêmement variables (Nakatani & Dukes, 1977) et aucun de ces indices ne peut constituer un marqueur de frontière lexicale fiable et stable (Klatt, 1980). La place et la fonction des pauses non physiologiques ont été également étudiées. S'il a été montré qu'elles étaient utilisées dans la plupart des langues comme marqueurs de frontière, elles se réalisent le plus souvent sur des frontières syntaxiques importantes (Vaissière, 1983 ; Grosjean et Deschamps, 1973, 1975) et permettent de dégager des constituants correspondant davantage à des unités de la taille du syntagme.

L'information (segmentale et suprasegmentale) qui est susceptible de signaler l'existence de frontières lexicales varie selon les langues. Les hypothèses de segmentation prélexicale ne peuvent donc être avancées que grâce à la multiplication de recherches inter-langues et doivent être mises en rapport avec la structure phonologique et prosodique de la langue envisagée. Deux types d'hypothèses de segmentation prélexicale ont été à ce jour envisagées : une hypothèse métrique et une hypothèse prosodique. Nous développons ici principalement les hypothèses métriques et prosodiques pour le français.

### **3.1. L'hypothèse métrique : les problèmes du français**

Les hypothèses de segmentation prélexicale ont d'abord été élaborées pour l'anglais. Cutler et Carter (1987) ont observé sur une large base de données que 74 % des voyelles pleines et accentuées se trouvaient en position initiale de mots de contenu. S'appuyant sur la régularité de cet indice de frontière dans le signal, Taft (1984), Cutler et Butterfield (1992), Cutler et Norris (1988) ont proposé une stratégie de segmentation dite « métrique » qui s'appuierait sur la succession des voyelles accentuées et pleines et des voyelles non accentuées et réduites.

Considérant que le français est une langue à accent fixe dont l'unité métrique est la syllabe <sup>7</sup>, Cutler, Otake, Hatano et Mehler (1993) ont proposé l'hypothèse syllabique en français comme variante de l'hypothèse métrique pour l'anglais.

La convergence des résultats obtenus en anglais, en japonais et en français a en effet amené les auteurs (Mehler, Dommergues, et Segui, 1981 ; Cutler, Mehler, Norris et Segui, 1983, 1986, 1987) à proposer l'hypothèse syllabique pour le français, comme une variante de segmentation contrainte par la structure de la langue. Cutler, Otake, Hatano et Mehler (1993) suggèrent que les auditeurs segmentent en utilisant une routine de segmentation mise en place lors de l'acquisition du langage et adaptée à l'unité métrique de leur langue. Cette conception les amène à proposer que les francophones segmentent la parole en

syllabes comme les anglais en pieds ou les japonais en morae. L'hypothèse sous-jacente résulte d'une conception en principes et paramètres : les enfants disposeraient de manière innée d'un principe universel de segmentation qui serait orienté ensuite en fonction de contraintes portant sur le choix de l'unité et de la stratégie pertinentes.

Une routine de segmentation strictement ascendante qui procéderait syllabe par syllabe ne peut pourtant pas être envisagée. Une conception strictement syllabique de la segmentation s'appuyant sur la syllabe relève d'une analogie post-hoc qui n'a pas été expérimentalement validée et qui reconduit une conception strictement linéaire de la segmentation. Or d'une part, les résultats de Mehler *et al.* (1981) concernent la reconnaissance d'une structure syllabique à l'intérieur d'un mot. Les auteurs ont montré que la syllabe est une unité de structuration, d'organisation à l'intérieur d'un mot et non une unité de segmentation. D'autre part, ces recherches s'appuient sur une représentation plate et linéaire de la syllabe dont le statut théorique n'est pas clairement posé. La chaîne est apparemment envisagée comme une succession de syllabes isomorphes et isochrones. Or une telle représentation syllabique de la parole ne se justifie pas davantage que la conception phonémique précédemment envisagée. Cette hypothèse de segmentation ainsi formulée n'est évidemment pas soutenable au regard des phénomènes de liaison enchaînée et d'enchaînement. Si les francophones développaient une stratégie de segmentation alignée sur les frontières syllabiques et considéraient que toutes les frontières de syllabes peuvent constituer un début de mot, ils segmenteraient systématiquement de manière erronée les suites contenant une liaison ou un enchaînement. Zwitserlood, Schriefers, Lahiri et Van Donselaar (1993) ont mené une recherche identique à celle de Mehler, Dommergues et Segui (1981) en néerlandais, qui est une langue à accent fixe comportant des frontières ambisyllabiques et des frontières claires. Ils proposent alors une stratégie syllabique prélexicale dynamique où les sujets ne concaténeraient pas des syllabes préstockées mais dériveraient la syllabation du continuum en temps réel :

Syllables are not represented units but rather derived or computed units [...], syllable recognition then does not exclusively depend on the availability of acoustic cues signalling syllable boundaries, but rather, it depends on the interplay between properties of the acoustic input and the listeners' knowledge about the phonological structure of their language. (p. 270)

Cette conception suppose donc un niveau intermédiaire qui permettrait de procéder en temps réel au calcul de l'organisation syllabique du continuum sonore. Une telle conception mérite d'être envisagée pour le français puisque, nous l'avons vu, ce sont essentiellement les phénomènes de resyllabation sur les frontières de mots à l'intérieur d'un domaine qui posent problème. Elle paraît être une hypothèse de segmentation prélexicale plausible qui, en envisageant un calcul en temps réel de la syllabation, permettrait de rendre compte de la liaison

enchaînée. Il faudrait alors spécifier quelle est la nature de ce niveau de représentation intermédiaire sur lequel se calcule la syllabation et envisager une segmentation prosodique incluant un calcul syllabique et supposant que les auditeurs utilisent une unité plus grande que le mot pour accéder au mot. Il convient alors d'envisager qu'ils usent de procédures rétroactives et ne traitent pas linéairement le signal.

### 3.2. La segmentation prosodique

Dans ce cadre, l'accès au lexique se produit sur la base d'une portion de signal prosodiquement contrainte, et plus grande que le mot. Les auditeurs utilisent donc une représentation prosodique et organisée du signal à partir de laquelle ils recherchent les frontières lexicales. Une telle conception a été envisagée dès 1987 par Grosjean et Gee, qui ayant souligné l'importance démesurée accordée au mot orthographique dans les modèles de traitement de la parole continue, ont proposé une segmentation prosodique du signal. Pour le français, elle a été ensuite développée par Christophe (1993), Christophe et Dupoux (1995) et Wauquier-Gravelines (1996).

Deux problèmes se posent dès lors qu'on envisage cette conception : quelles sont la taille et la nature de l'unité prosodique permettant l'accès au lexique ?

Est-ce que cette unité constitue un domaine de traitement et que la segmentation suppose alors des phénomènes rétroactifs contrariant la représentation linéaire que nous avons et du signe et du signal ?

#### 3.2.1. Nature et taille de l'unité de traitement

Concernant la nature et la taille de l'unité de segmentation prosodique, plusieurs solutions ont été proposées. De nombreux psycholinguistes se sont intéressés à l'existence d'unités prosodiques de traitement en interrogeant le décalage entre les représentations abstraites et syntaxiquement contraintes proposées par la théorie linguistique, et les unités obtenues par les procédures de découpages de l'énoncé. Dans un premier temps, des recherches ont, en particulier, tenté de tester la réalité psychologique des structures de surface proposées par la grammaire générative et transformationnelle (Levelt, 1969 ; Johnson, 1965, 1969), sans jamais parvenir à établir une correspondance stricte avec les constituants syntaxiques proposés par la grammaire générative.

D'autres recherches ont proposé une description relevant de la performance et se sont appuyées sur les informations acoustiques présentes dans le signal pour déterminer des unités de traitement (Ferreira, 1993 ; Gee et Grosjean, 1983 ; Grosjean et Dommergues, 1983 ; Monnin et Grosjean, 1993 ; Martin, 1970 ; Delais, Padeloup). Les corrélats acoustiques suivants ont été retenus : *le patron intonatif* : le ou les niveau(x) intermédiaire(s) se caractérise(nt) par la présence d'un patron montant-descendant, assurant une fonction démarcative ; *la distribution des accents* : les unités définies constituent les

domaines de réalisation des accents primaires et secondaires (Delais, 1994, 1995 ; Padeloup, 1989, 1990) ; *les indices de durée* (Gee & Grosjean, 1983 ; Grosjean et Dommergues, 1983 ; Monnin et Grosjean, 1993 ; Delais, 1994, 1995 ; Padeloup, 1989, 1990, Ferreira, 1993).

Toutes ces unités s'inscrivent dans une structure hiérarchique pouvant se résumer ainsi : on peut modéliser l'existence dans le signal d'un domaine minimal de structuration qui est la syllabe, d'un domaine maximal qui est la phrase ou l'énoncé, et d'un domaine intermédiaire que nous appelons ici, de manière neutre et générique, un groupe prosodique. Selon le cas, cette étape intermédiaire est constituée de un ou deux niveaux.

Dans tous les modèles, un principe d'équilibrage est respecté lors du découpage de la parole. Les locuteurs ont tendance à segmenter l'énoncé en vue de la préservation d'une symétrie entre les constituants à un niveau donné et à l'intérieur du niveau strictement supérieur. Ceci s'opère également parfois au détriment de la structure syntaxique. Dans tous les cas, la taille des constituants intermédiaires dans la structure est également contrainte par le nombre de syllabes qui les constituent. Ces unités intermédiaires dans la structure ne sont jamais composées d'une seule syllabe, et à plus forte raison d'une seule syllabe non accentuée. Par ailleurs, leur longueur moyenne est de 3 à 4 syllabes et n'excède jamais 6 ou 7 syllabes. Delais et Padeloup ont montré que le principe d'équilibrage et le principe de respect du nombre de syllabes interagissent. Il semblerait que le but visé par les locuteurs consiste à ne pas allonger les constituants au-delà d'une certaine limite, et ce, quelles que soient les caractéristiques métriques de la langue observée (anglais ou français).

Ce fait incite à penser que la limite concernant la taille de l'unité prosodique produite est imposée par la capacité cognitive d'encodage des informations. (Wauquier, 1996 ; Padeloup, 1989, 1990 ; Levelt, 1989, 1992.)

Les recherches présentées ont toutes montré que l'organisation du signal ne correspond pas à la structure syntaxique de surface. Pourtant, les points de vue divergent quant au degré d'isomorphisme que l'on peut établir entre les constituants produits et les unités prosodiques proposées par la hiérarchie prosodique, dérivée de la structure syntaxique (Nespor et Vogel, 1986 ; Selkirk, 1984).

Grosjean, Grosjean et Lane (1979), Gee et Grosjean (1983), Grosjean et Dommergues (1983) ont rapproché ces structures obtenues sur des phrases anglaises des unités de la première version de la hiérarchie prosodique proposée par Selkirk (1978, 1980). Ils suggèrent que le syntagme phonologique constitue une unité de traitement. Grosjean, Grosjean et Lane (1979) ont montré en particulier que 74 % des unités de base déterminées par leur étude des pauses correspondaient à des syntagmes phonologiques. Les auteurs se montrent beaucoup plus réservés dans une version française des structures de performance (Monnin et Grosjean, 1993) quant à la correspondance entre structures de performance et constituants de la phonologie prosodique. Le parallèle entre ces deux types de constituants n'est pas évoqué.

Ferreira considère à l'inverse que les marques de durée sont la trace dans le signal d'une contrainte prosodique. Selon l'auteur, ces marques de durée sont les indices de la mise en œuvre au niveau du signal de la structure prosodique dont elle s'inspire (Selkirk, 1984). Elle n'envisage pas, comme l'ont fait Grosjean *et al.*, qu'il existe un niveau de représentation phonétique ayant ses propres caractéristiques, contraint par des informations de performance et parallèle à la représentation phonologique.

Pour conclure, toutes ces recherches permettent de faire l'hypothèse de l'existence, en production, d'une unité intermédiaire entre la syllabe et la phrase, constituée de 4 syllabes en moyenne, dont la longueur est limitée à 6 ou 7 syllabes et tend à être préservée autant que possible par le locuteur. Elle est associée à un patron prosodique montant-descendant, ainsi qu'à un patron accentuel (pour le français : syllabe accentuée précédée de une ou plusieurs syllabes non accentuées).

### 3.2.2. Unité de perception

Les auditeurs peuvent-ils se servir de ce type d'unités pour segmenter la parole continue et accéder au lexique ? Cette unité de production est-elle susceptible d'être utilisée en perception ?

Nous n'avons effectivement pour l'instant, aucune certitude quant à la symétrie des processus de production et de perception. Mais aucun argument dirimant n'invalide non plus l'hypothèse de la symétrie entre procédures perceptives et production. L'auditeur d'une langue en est aussi un locuteur. Les groupes prosodiques que le locuteur a produit et dans lesquels il a opéré des opérations phonologiques de réajustement sont repérables par des indices suprasegmentaux. Ils correspondent le plus souvent à de grandes articulations syntaxiques ou argumentales et n'excèdent jamais une certaine durée. De plus le locuteur les produit en vue d'établir une communication efficace avec son interlocuteur et les rend perceptibles.

On peut faire l'hypothèse que ces informations convergentes proposent à l'auditeur un domaine de traitement à partir duquel il peut retrouver des représentations lexicales. Grosjean & Gee (1987) ont proposé que la segmentation se fasse à partir d'une unité prosodique. Christophe et Dupoux (1995) ainsi que Wauquier-Gravelines (1996) ont produit des modélisations de segmentation prosodique pour le français s'appuyant sur des résultats expérimentaux. Dans tous les cas, il est proposé une première segmentation de la chaîne en constituants prosodiques dits « domaines perceptifs » (Wauquier-Gravelines, 1996) ou « structures de performances » (Grosjean *et al.*), constituants définis exclusivement par des contraintes de performance et des contraintes phonétiques (accents, allongements, patrons intonatifs). C'est à ce niveau qu'est traitée l'information acoustique concernant la structure métrique de l'entrée parlée. En français par exemple c'est à ce niveau que serait déterminée la structure syllabique de la chaîne.



L'accès au lexique est réalisé par activation parallèle de l'information segmentale et suprasegmentale et par convergence de cette double information avec l'information lexicale descendante multilinéaire phonologiquement représentée. Cette double dynamique permet que la sélection des mots s'opère au niveau des constituants prosodiques abstraits après traitement des divers phénomènes phonologiques (sandhis, resyllabations) sur leur domaine de réalisation. Ces constituants relèvent d'un niveau abstrait de représentation, appartenant à la compétence phonologique que les locuteurs utilisent pour traiter la parole continue. De ce point de vue, la phrase intonationnelle et l'énoncé prosodique nous semblent des unités peu intéressantes. Elles sont longues, difficilement stockables en mémoire à court terme et nous reconduisent aux problèmes précédemment envisagés concernant la segmentation du continuum : si la reconnaissance des mots doit s'effectuer sur des unités très longues et syntaxiquement complexes, on retrouve le problème du traitement d'une phrase complexe continue constituée d'un nombre important de mots appartenant à différentes classes morphologiques. Le groupe clitique et le syntagme phonologique par leur taille et leur cohésion syntaxique constituent les domaines prosodiques à l'intérieur desquels pourrait s'opérer l'accès lexical. Ce sont les unités les plus petites sur lesquelles s'appliquent des phénomènes phonologiques entre les mots comme la liaison enchaînée en français (De Jong, 1991) ou d'autres phénomènes de sandhis.

Le syntagme phonologique est en outre le domaine de la resyllabation en français, en anglais, en grec (Nespor et Vogel, 1986). Le syntagme phonologique pourrait alors constituer le niveau de représentation intermédiaire, mentionné par Zwitserlood *et al.*, Grosjean et Gee, où se calcule la syllabation de la chaîne permettant de segmenter le signal. Dans le cas de la liaison par exemple (Wauquier-Gravelines, 1996), c'est à ce niveau que pourrait se résoudre le décalage entre la syllabation de surface [pø/ti/ta/mi] et la structure syllabique des unités lexicales. Les résultats expérimentaux obtenus concernant le traitement de chaînes ambiguës (un grand téléphone/un grand éléphant) [Wauquier-Gravelines, 1996] montrent que les auditeurs possèdent dans leur lexique une représentation de « grand » comme /grāt/ et qu'ils l'activent préférentiellement sur un format prosodique tel que le syntagme phonologique. Racine et Grosjean (à paraître) envisagent également ce type de réponse au problème du traitement des frontières lexicales dans le cas d'effacement du schwa. De tels modèles envisagent donc qu'il y ait lors des procédures de traitement un ajustement permanent entre l'information perceptivement disponible et représentations phonologiques abstraites multilinéaires.

### 3.3. Pour en revenir à la question de la linéarité

Les modèles précédemment évoqués, qu'ils soient métriques ou prosodiques, nous amènent à reposer la question de la linéarité. Ils ne présupposent pas une représentation abstraite de la substance projetée au niveau

du signal et correspondant à ce titre à des traces acoustiques. Comme nous venons de le voir, ces modèles s'appuient prioritairement sur la linéarité du signal et la présence d'une information prosodique récurrente. Le traitement linéaire est donc bien réalisé mais sur la base d'une information physique qui permet de dégager des unités plus grandes que le mot. Ceci signifie que le signal de parole est bien un objet linéaire qui peut être décomposé en une suite de groupes prosodiques physiquement marqués mais que les unités lexicales vont être ensuite retrouvées à partir de ces unités prosodiques. À ce traitement linéaire sont donc associées des possibilités de procédures rétroactives permettant le calcul des frontières lexicales à l'intérieur du domaine prosodique pertinent. Ne procédant pas de manière strictement linéaire, les modèles prélexicaux doivent inclure un stockage temporaire des portions de signal ainsi découpées, en attente de segmentation lexicale. Ce stockage ne peut pourtant pas être envisagé comme une réserve provisoire et inerte de signal en attente de traitement. L'information acoustique étant délivrée de manière continue, les manœuvres rétroactives doivent être rapides, et extrêmement efficaces. Il est plus probable d'envisager que la segmentation prosodique ou métrique relève d'un processus actif où le stockage de l'information s'opère de manière heuristique et inférentielle. L'auditeur stockant une portion de signal acoustiquement pertinente effectuerait à ce moment là une première série d'hypothèses concernant 1) la nature du constituant entendu 2) la nature des opérations phonologiques (liaisons, élisions), morphologiques (dérivations, flexions) et syntaxiques (remontée d'un constituant) possibles sur ce domaine. Le coût cognitif qu'impose la procédure rétroactive serait alors compensé par la probabilité des découpages sur une unité donnée et l'élimination de certaines segmentations nécessairement non pertinentes. La procédure rétroactive n'interviendrait que pour les cas marqués quand l'inférence s'est trouvée contredite par l'information physique.

Si l'on reprend l'exemple des mots enchâssés contenus dans [lekrɔkɔdiltravɛrslɛmarigo], un auditeur francophone qui en stockant l'unité syntaxique et prosodique que constitue [lekrɔkɔdil] repère qu'il s'agit d'un GN sujet isolable par une série d'informations prosodiques, élimine automatiquement des segmentations lui permettant de trouver « croque, Odile, ode, île ». Dans le cas d'une liaison dans un syntagme phonologique, Wauquier-Gravelines a montré que les auditeurs n'ont pas eu plus de difficulté à trouver le mot « éléphant » dans « un petit éléphant » que dans « un joli éléphant ». Au moment de l'apparition de la consonne de liaison, quand l'auditeur a reçu l'information acoustique [pøtit], il sait qu'il a une probabilité de liaison très élevée et qu'il a beaucoup plus de chance, dans un syntagme phonologique, d'être en présence de /pəti + (t)/ + /e.../ que d'être en présence de /pəti/ + /te.../. Il peut donc privilégier la première segmentation et faire l'hypothèse que ce segment /t/ ne constitue pas un segment initial de mot et entamer la recherche du mot suivant à partir de la voyelle avec une probabilité d'erreur très faible. Dans ce modèle de traitement, incluant une représentation lexicale du segment flottant c'est la distribution la plus fréquente (déterminant ou adjectif antéposé terminé par une consonne de liaison + voyelle,

par exemple : petit éléphant) qui est favorisée comme un cas plus probable de segmentation au détriment de la distribution nécessairement plus rare (déterminant ou adjectif antéposé terminé par une consonne de liaison + consonne initiale de mot identique, par exemple : petit téléphone). C'est cette seconde distribution qui constitue l'anomalie et nécessite une confirmation après l'accès lexical.

## Conclusion

Si les hypothèses de segmentation non strictement linéaires ont donné lieu à moins de recherches expérimentales et reposent donc sur moins de certitudes empiriques que les conceptions précédemment envisagées, elles ont définitivement invalidé la conception strictement linéaire et séquentielle du traitement de la parole continue. Il ne s'agit pas bien évidemment de nier l'existence de procédures linéaires: la continuité et la temporalité du signal de parole nécessitent qu'on modélise la séquentialité ; il convient d'associer aux procédures linéaires la possibilité d'opérations rétroactives permettant de rendre compte du calcul des frontières lexicales et morphémiques à partir de constituants prosodiquement déterminés. Une telle conception est théoriquement beaucoup plus défendable que les autres solutions apportées à la segmentation. (En ce qui concerne la validité psycholinguistique de cette hypothèse voir Wauquier-Gravelines, 1996.) Elle permet d'envisager que les auditeurs utilisent toute l'information contenue dans le signal et ne négligent pas en particulier l'information prosodique dont la fonction démarcative est apparemment universelle. D'autre part, elle intègre une utilisation active de la grammaire lors des procédures de traitement en temps réel. La variété des problèmes de segmentation d'un continuum sonore que peuvent poser les milliers de langues parlées sur la terre est telle que l'on imagine difficilement qu'une seule procédure algorithmique nécessitant le stockage d'un nombre considérable d'informations inutiles puisse rendre compte de la diversité des procédures effectivement utilisées. Il paraît beaucoup plus probable que les auditeurs procèdent de manière heuristique et se servent d'un système de traitement beaucoup plus adaptatif utilisant toute l'information dont ils disposent et donc, qu'ils contraignent ses procédures de traitement par l'utilisation de la compétence phonologique.

## NOTES

1. Opposition entre modèles modulaires dérivationnels et modèles en réseaux à activation et inhibition simultanée. Pour une synthèse critique sur ce sujet, voir Wauquier-Gravelines (1996), Mattys (1997).
2. « *One word's recognition automatically directs segmentation of the immediately following word.* » (Cole et Jakimik, 1980.)
3. Il a soumis aux sujets des séquences telles que « *I saw the bun in the oven* » (J'ai vu le petit pain au four). Le mot à reconnaître (*bun*) est suivi d'une préposition (*in*). Il a montré que le mot « *bun* » était reconnu en même temps voire après la préposition qui le suit et que des hypothèses

concurrentes étaient proposées par les sujets au fur et à mesure du dévoilement de l'information acoustique (*bun, bunny, etc.*).

4. Des recherches ont été faites par Harrigton et Johnstone (1987) sur le nombre d'hypothèses de segmentations exhaustives envisageables en anglais : ils ont trouvé que pour 115 phrases transcrites phonétiquement contenant en moyenne de quatre à six mots, mais jamais plus de dix mots, les possibilités de segmentation exhaustives (c'est-à-dire de segmentation où chaque phonème n'appartient qu'à un seul mot de la phrase) dépassent plus de 10 possibilités dans 65 % des cas. Frauenfelder (1991b) mentionne également une étude statistique effectuée sur environ 10 000 mots néerlandais isolés : 46% des mots peuvent constituer le début, la partie centrale ou la fin d'un autre mot plus long. Il signale que la possibilité de segmentation exhaustive est environ vingt fois plus élevée quand les mots apparaissent en parole continue. Il est évident que les resyllabations opérées sur les frontières lexicales multiplient les possibilités de segmentation.
5. Les sujets reçoivent un mot inséré dans une phrase en modalité auditive et doivent opérer une décision lexicale sur un mot présenté en modalité visuelle (sur un écran). Quand le mot présenté en modalité auditive et le mot sur lequel doit se faire la décision lexicale sont sémantiquement reliés, on constate une facilitation de la décision lexicale. Si les sujets doivent par exemple faire une décision lexicale sur le mot « chien » (indiquer si «chien » est un mot ou un non-mot) la performance est plus rapide quand le mot « chien » est précédé en modalité auditive de « animal», qui est un hyperonyme, que s'il est précédé de «addition » avec lequel il n'entretient aucune relation sémantique. Ceci laisse penser que le fait d'avoir entendu un mot apparenté a activé « chien » qui est ainsi reconnu plus vite.

## RÉFÉRENCES

- ASLIN, R.N. ; PISONI, D.B. ; USCZYK, P.W. (1983). Auditory development and speech perception in infancy. Dans Haith & Campos (eds.), *Carmichael's Handbook of Child Psychology : Infancy and Developmental Psychobiology*, New York : Wiley.
- BARD, E.G. ; HILLCOCK, R.C. ; ALTMANN, G. T. M. (1989). The recognition of words after their acoustic offsets : effect of subsequent context. *Perception and Psychophysics*, 44 : 395-408.
- BERTONCINI, J. (1994). Les capacités linguistiques du nouveau-né. Dans Aslanides-Rousselet & Wauquier-Gravelines (eds), *L'Enfance du Langage, Cahiers de Fontenay*, 75, Presses de l'École Normale Supérieure de Fontenay St-Cloud.
- CAO XUAN HAO (1985). *Phonologie et linéarité, réflexions critiques sur les postulats de la phonologie contemporaine*. Numéro spécial de la Société d'Études Linguistiques et Anthropologiques de France, 18, Paris.
- CHRISTOPHE, A. (1993). *Rôle de la prosodie dans la segmentation en mots*, Thèse soutenue en vue de l'obtention du Doctorat de l'EHESS, Paris.
- CHRISTOPHE, A. ; DUPOUX, E. (1995). Bootstrapping lexical acquisition : the role of prosodic structure. *The linguistic Review*.
- COLE, R.A. ; JAKIMIK, J. (1980). A model of Speech Perception. Dans Cole (ed.), *Perception and Production of Fluent Speech*, Hillsdale, NJ : Lawrence Erlbaum, 135-163.
- 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 the English vocabulary. *Computer Speech and Language*, 2: 133-142.
- CUTLER, A. ; MEHLER, J. ; NORRIS, D. ; SEGUL, J. (1986). The Syllable's Differing Role in the segmentation of French and English. *Journal of Memory and Language*, 25 : 385-400.

- CUTLER, A., MEHLER, J., NORRIS, D., SEGUI, J. (1987). Phoneme Identification and the lexicon. *Cognitive Psychology*, 19 : 141-177.
- CUTLER, A., NORRIS, D., SEGUI, J. (1983). A language specific comprehension strategy. *Nature*, 304 : 159-160.
- CUTLER, A., NORRIS, D., SEGUI, J. (1983). A language specific comprehension strategy. *Nature*, 304: 159-160
- DE JONG, D. (1991). La liaison à Orléans (France) et Montréal (Québec). *Actes du XII<sup>e</sup> Congrès International des Sciences Phonétiques*, août 1991, 198-201. Aix-en Provence, France.
- DELAIS, E. (1994). Structure rythmique et prosodique du français : pour une approche psychocognitive. *Linx*, 29, Université Paris X Nanterre.
- DELAIS, E. (1995). Pour une approche parallèle de la structure prosodique. Étude de l'organisation prosodique et rythmique de la phrase française. Thèse de Doctorat. Université de Toulouse-Le Mirail.
- FERREIRA, F. (1993). Creation of Prosody during Sentence Production. *Psychological Bulletin*, 100 : 233-253.
- FRAUENFELDER, U. ; KOMISARJEVSKY-TYLER, L., (1987). The process of spoken word recognition. *Cognition*, 25: 1-20.
- FRAUENFELDER, U.H. ; SEGUI, J. ; DIJKSTRA, T. (1990). Lexical effects in phonemic processing : Facilitory or inhibitory ? *Journal of Experimental Psychology : Human Perception and Performance*, 16: 77-91.
- GEE, J. P. ; GROSJEAN, F. (1983). Performance Structures : A Psycholinguistic and Linguistic Appraisal. *Cognitive Psychology*, 15 : 411-488.
- GOW, D. W. ; GORDON, P. C. (1995). Lexical and Prelexical Influences on Word Segmentation : Evidence from priming. *Journal of Experimental Psychology : Human Perception and Performance*, vol. 21, n° 2 : 344-359.
- GROSJEAN, F. (1980). Spoken word Recognition processes and the gating paradigm. *Perception & Psychophysics*, 28 : 267-283.
- GROSJEAN, F. (1985). The Recognition of words after their acoustic offset : evidence and implications. *Perception and Psychophysics*, 38 : 299-310.
- GROSJEAN, F. ; DESCHAMPS, A. (1973). Analyse des variables temporelles du français spontané. *Phonetica*, 28 : 191-226.
- GROSJEAN, F. ; DESCHAMPS, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français. *Phonetica*, 31, 144-184.
- GROSJEAN, F., (1985). The Recognition of words after their acoustic offset : evidence and implications. *Perception and Psychophysics*, 38, 299-310.
- GROSJEAN, F. ; GEE, J. P. (1987). Prosodic Structure and Spoken Word Recognition. Dans U. H. Frauenfelder, L. Komisarjewsky-Tyler (eds.), *Spoken Word Recognition : Cognition, Special Issue*, Vol. 25 : 135-157.
- KLATT, D. H. (1980). Problem of variability in speech recognition and models of speech perception. Dans Perkell, Klatt (eds.), *Invariance and variability in speech processes*, Hillsdale, N.J. : Erlbaum.
- LAHIRI, A. ; MARSLÉN-WILSON, W. D. (1991). The mental Representation of Lexical Form : A phonological approach to the recognition lexicon. *Cognition*, 38: 245-294.
- LAKS, B. (1996). *Langage et cognition*, Paris : Hermès.
- LECOCQ, P. ; SEGUI, J. (1989). L'accès lexical. *Lexique*, 8 : 7-12
- LEHISTE, I. (1965). The function of quantity in Finnish and Estonian. *Language*, 41 : 447-456.
- LEHISTE, I. (1966). Consonant quantity and phonological units in Estonian. *Indiana University Publications of the Uralic and Altaic Series*, 65, Mouton : The Hague.
- LEVELT, W. J. M. (1989). *Speaking. From Intention to Articulation*, Cambridge: Bradford Book.

- LEVELT, W. J. M. (1992). Accessing words in speech production : stages, processes, and representations. *Cognition* 42 : 1-22.
- LEVELT, W. J. M. (1993). Timing in Speech production with special reference to word form encoding. Dans Tallal, gallaburda, Livas et Euler (eds.), *Temporal Information processing in the nervous system*, 283-295. Annals of the New York Academy of Science, vol. 582.
- LIBERMAN, A. M. (1970). Some characteristics of perception in the speech mode. Dans P. Hamburg, K. Pribram & A. J. Standard (eds.) *Perception and its disorders*, 238-254. Baltimore : Williams & Wilkins.
- LUCE, P. A. (1986). *Neighborhoods of words in the mental lexicon*, Doctoral Dissertation, Indiana University, Bloomington, Indiana.
- MARSLÉN-WILSON, W. D. (1987). Fonctionnel parallelism in spoken-word recognition. Dans Frauenfelder, Komisarjewsky-Tyler (eds.), *Spoken Word Recognition*, 72-102. *Cognition, Special Issue*, Vol.25.
- MARSLÉN-WILSON, W. D. ; WARREN, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes and features. *Psychological review*, vol. 101: 653-675.
- MARSLÉN-WILSON, W. D. ; WELSH, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10 : 29-63.
- MATTHEWS, P. H. (1974) : *Morphology*, Cambridge Text books in Linguistics.
- MATTYS, S. (1997). The use of time during lexical processing and segmentation : A review. *Psychonomic Bulletin & Review*, 4 : 310-329.
- MC CLELLAND, J. L. ; ELMAN, J. L. (1986). Interactive Processes in Speech Perception : The Trace Model. Dans Mc Clelland, Rumelhart, PDP Research Group (eds.), *Parallel Distributed Processing*, Vol 2, 58-121. Psychological and Biological Models, Cambridge, Mass : The MIT Press.
- MEHLER, J. ; DOMMERGUES, J. Y. ; SEGUI, J. (1981). The syllable's Role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20 : 298-305.
- MEHLER, J. ; DUPOUX, E. ; SEGUI, J. (1990). Constraining Models of lexical Access: the onset of word recognition. Dans Altman (ed.), *Cognitive models of Speech Processing : Psycholinguistic and Computational Perspectives*, 263-280. Cambridge, Mass : The MIT Press.
- MILNER, J.-C. (1989). *Introduction à une science du langage*, Paris: Le Seuil.
- MONNIN, P. ; GROSJEAN, F. (1993). Les structures de performance en français : caractérisation et prédiction. *L'Année Psychologique*, 93 : 9-30.
- NAKATANI, L. H. ; DUKES, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, 62 : 714-719.
- NESPOR, M. ; VOGEL, I. (1986). *Prosodic Phonology*, Dordrecht : Foris.
- OLLER, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, 54: 1235-1247.
- PASDELOUP, V. (1989). Essai de modélisation acoustique du rythme du français. *Journal d'acoustique*, 2-1 : 47-51.
- PASDELOUP, V. (1990). Modèle de règles rythmiques du français appliqué à la synthèse de la parole. Thèse de 3<sup>e</sup> cycle, Université de Provence, Aix- Marseille I.
- QUÉNÉ, H. (1992). Durational cues for word segmentation in dutch. *Journal of Phonetics*, 20 : 331-350.
- RACINE, I. ; GROSJEAN, F. (à paraître). Effacement du schwa, frontière lexicale et reconnaissance des mots en parole continue.
- RADEAUN, M. ; MOUSTY, P. ; PASDELOUP, V. (1994). L'effet du débit de la parole sur le point d'unicité. *Actes du colloque des XX<sup>e</sup> journées sur la Parole*, Trégastel.
- SELKIRK, E. (1984). *Phonology and syntax : the relation between sound and structure*, Cambridge, Mass.: MIT Press.

- SELKIRK, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook* 3 : 371-405.
- SPENCER (1991). A. *Morphological Theory, an introduction to Word Structure in Generative Grammar*; Blackwell.
- STUDDERT-KENNEDY, M. (1975). Speech Perception. Dans N. J. Lass (ed.), *Contemporary issues in experimental phonetics*, 243-293. New-York : Academic Press.
- TAFT, L. A. (1984). *Prosodic constraints and lexical parsing strategies*. Unpublished PhD thesis, University of Massachusetts at Amherst.
- TAFT, M. ; HAMBLY, G. (1986). Exploring the Cohort Model of spoken word recognition. *Cognition*, 22: 259-282.
- UMEDA, N. (1975). Vowel Duration in American English. *Journal of Acoustical Society of America*, 58 : 434-445.
- UMEDA, N. (1977). Consonant Duration in American English. *Journal of Acoustical Society of America*, 61 : 846-858.
- VAISSIÈRE, J. (1995a). Natural explanations for prosodic cross-languages similarities. *Proceedings of XIIIth Congress of Phonetic Sciences*, Stockholm, vol. 4 : 654-657.
- VAISSIÈRE, J. (1995b). Phonetic explanations for Cross-Linguistic Prosodic Similarities. *Phonetica*, 52 : 123-130.
- VAISSIÈRE, J. (1983). Language-independent prosodic features. Dans Cutler & Ladd (eds.), *Prosody, models and measurements, Springer Series in Language and Communication*, 14: 53-66. Berlin, Heidelberg, New-York, Tokyo : Springer-Verlag.
- 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.
- WAQUIER-GRAVELINES, S. (1996). *Organisation phonologique et traitement de la parole continue, contraintes prosodiques et phonologiques de l'accès lexical*, Thèse de doctorat de linguistique théorique et formelle, Université Paris 7, non publiée.
- ZWITSERLOOD, P. ; SCHRIEFERS, H. ; LAHIRI, A. ; VAN DONSELAAR, W. (1993). The role of Syllables in the Perception of Spoken Dutch. *Journal of Experimental Psychology : Learning Memory and Cognition*, vol. 19, n° 2 : 260-271.

#### ABSTRACT

Speech signal and phonological representations of speech adopted by psycholinguistical models (SPE model) have supported for a long time a strictly left-right algorithmic linear conception of speech treatment. Last experimental results and last psycholinguistical modelisations, as also proposals of multilinear phonology, invite us to consider that speech treatment could be more heuristic and could include both proactive and retroactive processes.

#### KEY-WORDS

Segmentation, linearity, speech treatment, phonology.

