Chapitre 2

Segmentation du signal de parole en mots

SEGMENTATION DU SIGNAL DE PAROLE EN MOTS

Il peut sembler étonnant de poser le problème des procédures mises en œuvre pour découvrir les frontières qui séparent les mots dans un message parlé comme un phénomène qui serait en partie indépendant des processus de reconnaissance des mots. Supposons en effet que les difficultés concernant la variabilité -spectrale et temporelle- du signal de parole aient été résolues, et que l'on soit aujourd'hui en mesure de proposer un modèle valide et non-réfuté de l'appariement entre image auditive et représentations phonétiques. Il serait alors possible de mettre au point un système artificiel capable d'identifier avec certitude la chaîne de phonèmes correspondant au signal acoustique qui lui est fourni en entrée. Supposons maintenant que ce système dispose également de procédures lui permettant de traduire la chaîne de phonèmes en représentations phonologiques. Ce système disposerait alors de représentations sous-jacentes directement appariables avec les éléments stockés dans son lexique. On peut penser qu'il suffirait alors d'envisager, à partir de cette chaîne d'entités phonologiques sous-jacentes, l'ensemble des suites de mots possibles pour retrouver la séquence de mots effectivement produite par le locuteur. De manière simplifiée, c'est l'une des stratégies qui ont été proposées dans plusieurs modèles de la segmentation lexicale en psycholinguistique. Cette procédure est aussi largement utilisée en Reconnaissance Automatique de la Parole (RAP). La segmentation du signal de parole en mots se fait alors par comparaison de la séquence phonémique avec les mots

qui sont stockés dans le lexique. La séquence la plus plausible est sélectionnée ; ce qui permet d'aboutir à une segmentation relativement efficace du signal de parole en mots. Nous verrons, dans ce chapitre, que cette stratégie n'est peut-être pas la mieux adaptée aux processus de traitement de la parole. Pour ne mentionner qu'un argument, on notera le rôle éventuel des connaissances de haut niveau dans les processus d'identification phonémique, notamment le rôle possible de l'identification lexicale dans l'appariement entre image auditive et représentations linguistiques. Si les connaissances lexicales ont un rôle à jouer dans l'identification phonémique, il est important d'étudier la segmentation en mots comme un problème en soi ; des procédures aptes à guider la localisation des frontières lexicales pourraient ainsi guider l'appariement entre représentation auditive du signal acoustique et formes phonologiques abstraites en facilitant la reconnaissance des mots. L'objet de ce chapitre est à la fois de présenter divers modèles de la segmentation du signal de parole en mots et -parallèlement- d'illustrer les difficultés spécifiques que pose l'étude de ces processus. Nous décrivons en premier lieu les modèles qui reposent sur des processus de sélection ou de compétition lexicale pour présenter des positions alternatives dans lesquelles est intégrée l'utilisation d'informations non-lexicales (métriques, phonotactiques, statistiques).

1. La segmentation comme conséquence de l'accès au lexique

Depuis quelques années, on envisage que l'une des sources essentielles d'information pour la localisation des frontières de mots dans un signal de parole est l'identification même des mots qui le constituent. Certains modèles d'accès au lexique se sont alors présentés comme un choix particulièrement bien adapté au problème de la segmentation lexicale d'un signal de parole continue. Ces modèles reposent essentiellement sur le principe de la *sélection progressive des candidats lexicaux* les mieux appariés avec l'entrée acoustique.

1.1. Cohort

Le modèle COHORT (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978) est le premier à introduire la notion de reconnaissance d'un mot (ou d'accès au lexique) comme un processus de sélection progressive parmi un ensemble de candidats lexicaux possibles. Ce modèle comporte 2 étapes : (1) une étape d'*activation* des candidats lexicaux appariés avec le début du signal suivie de (2) une étape de *dé-activation* des candidats dont la structure phonétique dévie du signal. Il repose sur l'une des caractéristiques essentielles du signal de parole : son déroulement dans le temps. Contrairement à l'identification de mots écrits, situation

dans laquelle le système visuel est en mesure de coder intégralement -pour des mots de longueur moyenne- l'information pertinente avec une seule fixation visuelle, la reconnaissance des mots parlés implique un *codage progressif* de l'information. Marslen-Wilson & Welsch (1978) utilisent cette particularité des stimuli acoustiques pour proposer un modèle spécifique de la reconnaissance des mots dans la modalité auditive. La première étape consiste à déclencher l'activation d'une *cohorte* de mots possibles à partir de l'information disponible au début du signal acoustique (les 150 premières millisecondes, ce qui correspond approximativement à un ou deux phonèmes). A partir de cet ensemble de mots possibles, le système exclut les candidats (dans COHORT I) ou réduit progressivement leur activation (dans COHORT II) quand ils ne sont plus appariés avec l'entrée acoustique. Ce processus se poursuit jusqu'à aboutir à un choix lexical unique. Le point qui correspond à l'instant auquel le signal acoustique fournit une information qui permet d'identifier sans ambiguïté le mot prononcé est appelé *Point d'Unicité* (PU) ou *Point de Reconnaissance* (PR)[11]. Nous ne nous attarderons pas ici sur les problèmes liés aux conditions d'exclusion des candidats de la cohorte (doit-on supprimer les candidats qui ne sont pas appariés avec l'entrée acoustique ou simplement diminuer leur niveau d'activation ? -COHORT I vs. COHORT II), pas plus que sur la question de l'existence de processus d'inhibition ascendante ou de rétroaction du lexique sur les représentations phonémiques. On consultera notamment Frauenfelder (1996) pour des discussions relatives à ces problématiques. Notre objectif est ici de définir les caractéristiques fondamentales du modèle en rapport avec le problème de la segmentation d'un signal de parole continue en mots afin de fournir les bases d'une analyse des données présentées comme reflétant le recours à des connaissances sur les régularités phonotactiques de la langue.

1.1.1. Prédiction des frontières lexicales avant la fin acoustique du mot

Le modèle COHORT (I, Marslen-Wilson & Welsh, 1978; ou II, Marslen-Wilson, 1987) apporte une solution au problème de la localisation des frontières lexicales dans un signal de parole continue par le biais du processus même d'accès au lexique. Pour cela, il permet d'identifier un mot avant sa fin acoustique. Or, si l'on est en mesure de reconnaître un mot avant qu'il ait été traité dans son intégralité, on est en mesure de savoir à l'avance où il va se terminer et, par conséquent, où commencera le mot suivant. Ainsi, le verbe /apɛrsəvwar/ ('apercevoir') peut être isolé comme unique dans le lexique français dès l'occurrence du phonème /ə/.

[11] Selon que l'on considère le point à partir duquel le mot est unique dans un dictionnaire de la langue (PU) ou celui à partir duquel les locuteurs l'identifient avec certitude (PR ; ces deux 'points' peuvent notamment différer en raison de la fréquence d'usage des mots).

Lorsqu'en modalité auditive on a traité le flux acoustique qui correspond à la séquence de phonèmes /apɛrsə/, il n'existe pas dans le lexique d'autre possibilité que ce verbe. On est donc en mesure, dès cet instant là, de prédire que la fin du mot en cours correspondra à la séquence /vwar/ et que ce mot se terminera avec le phonème /r/. On peut alors prévoir avec certitude que le mot suivant commencera immédiatement après le phonème /r/. Il est par conséquent possible, grâce à des processus de reconnaissance des mots, de localiser les frontières lexicales avant qu'elles apparaissent effectivement.

L'aptitude du modèle COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987) à déclencher le processus d'accès lexical en parole continue repose essentiellement sur sa capacité à localiser le début des mots -on parle de modèle à *alignement initial*. En effet, c'est l'information acoustique fournie par le début du signal acoustique qui permet d'engendrer un ensemble d'hypothèses lexicales sur lesquelles vont s'opérer les processus de sélection. Seuls les mots dont le début est apparié avec l'entrée acoustique sont intégrés dans la cohorte. Ce modèle nécessite donc, dans le cas du traitement d'un signal de parole continue, une prédiction correcte de la prochaine frontière lexicale pour être en mesure de générer une cohorte pertinente pour le mot suivant. Si le système omet une frontière lexicale, il n'est plus en mesure de générer en temps réel la cohorte initiale de candidats lexicaux appariés avec l'entrée. Tout retard rencontré dans la localisation de la frontière lexicale induit alors une impossibilité -au moins transitoire- de générer une cohorte correspondant au début du mot suivant. Dans cette situation, seule une procédure de récupération par un retour en arrière dans les processus de traitement permettrait de récupérer le cours normal du traitement. Il n'est pas envisagé de pouvoir générer un ensemble de candidats qui seraient appariés avec le signal autrement que par leur début. Ce modèle est donc efficace uniquement s'il existe dans le lexique une grande quantité de mots -sinon tous- qui peuvent effectivement être reconnus avant leur fin acoustique. Supposons que l'on traite par exemple la suite /elef/, cette séquence de phonèmes permet de restreindre la cohorte à quelques mots du lexique français : 'éléphant', 'éléphanteau', 'éléphantiasis'. L'existence de ces deux derniers empêche le système d'accès au lexique d'être en mesure d'affirmer -dès l'occurrence du phonème /f/- que le prochain mot dans la séquence de parole commencera après le son /ɑ̃/. Il pourrait tout aussi bien commencer après le son /o/ si le mot en cours de traitement est en réalité 'éléphanteau'. On objectera que ceci ne met pas en échec le processus de localisation des frontières lexicales à partir de l'accès au lexique. Effectivement, si le mot devient unique sur son dernier phonème, il permet d'insérer une frontière lexicale et de relancer un nouveau processus d'accès au lexique à partir de ce point. Cependant, cette objection n'est valable que si c'est

effectivement le mot ‘éléphanteau’ qui est présenté. Si le mot à identifier est en fait ‘éléphant’ et que la séquence de phonèmes subséquente commence par /g/ comme dans ‘un éléphant gris’, c’est l’occurrence du phonème /g/ qui conduira le système à identifier le mot ‘éléphant’ (puisqu’il n’y a pas dans le lexique de mot commençant par /elefɑ̃g/). Le système devra alors revenir sur ses prédictions pour générer la cohorte pertinente à partir du son /g/. COHORT II (Marslen-Wilson, 1987) apporte en partie une solution à ce problème en introduisant la notion de niveau d’activation et en faisant appel à la fréquence des mots pour moduler l’activation des candidats. Le processus de sélection de la réponse appropriée (qui est alors assimilable à un processus intégratif) prend appui sur la fréquence relative des différents candidats activés et favorise les plus fréquents. Dans l’exemple précédent, ‘éléphant’ est beaucoup plus fréquent que les mots ‘éléphanteau’ et ‘éléphantiasis’. Il aura donc plus tendance à se maintenir dans la cohorte ; ce qui évitera peut-être un rejet trop précoce si le seuil de reconnaissance est bien fixé. On peut alors envisager que le modèle déclenche l’activation d’une cohorte sur le phonème /g/ avant même de pourvoir prédire que c’est effectivement le bon choix avec 100% de certitude.

1.1.2. Point d’Unicité tardif et génération d’une cohorte

Les difficultés inhérentes à ce modèle s’accentuent donc s’il existe dans le lexique des mots qui, même au niveau de leur fin acoustique, ne sont toujours pas uniques. Dans l’analyse d’un lexique anglais informatisé (Kucera & Francis, 1967), Luce (1986) montre que la quantité des mots qui ne sont pas uniques sur leur phonème final est considérable (74% des mots de 3 phonèmes et 36% des mots de 4 phonèmes ; cf. Tableau 1). Ce phénomène est d’autant plus manifeste que les mots sont courts. Frauenfelder & Peeters (1990), dans une analyse similaire effectuée sur un échantillon de mots néerlandais provenant de la base de données CELEX, obtiennent des données tout à fait comparables (61% de mots à PU tardif pour 3 phonèmes, 31% pour 4 phonèmes, 15% pour 5 phonèmes). Une stratégie consistant à prédire les frontières lexicales par le biais de processus de sélection reposant sur un alignement initial ne serait donc réellement efficace que pour le traitement des mots longs. Dans le cas des mots courts, le système devrait soit postuler des frontières lexicales en permanence afin de ne pas en omettre une seule, soit mettre en place un processus de retour en arrière lorsqu’il s’aperçoit qu’il en a laissé passer une (ceci même s’il incorpore un paramètre fréquentiel pour maintenir les mots fréquents -qui sont aussi les plus courts- dans la cohorte initiale).

Tableau 1 : Pourcentage de mots dans le lexique anglais dont le Point d'Unicité se situe après le phonème final. Extrait des résultats de Luce (1986) pour les mots comptant entre 3 et 7 phonèmes.

Nb. de phonèmes	Nb. de mots dans l'échantillon	Pourcentage de mots
3	1839	74.17
4	3025	35.64
5	3172	16.36
6	3063	8.56
7	2735	5.59
8	2210	4.75

L'existence d'un délai entre la fin acoustique des mots à PU tardif et leur reconnaissance est mise en évidence par Grosjean (1985). Il utilise une tâche de dévoilement progressif (*Gating*) dans laquelle des auditeurs écoutent des phrases contenant une quantité d'information acoustique variable. Les stimuli que reçoivent les participants vont progressivement des premières périodes acoustiques à l'intégralité du signal (cf. Figure 6 pour un exemple des stimuli utilisés dans une tâche de gating). Grosjean (1985) montre que les mots courts ayant des prolongements possibles dans le lexique ne sont identifiés avec un taux de certitude élevé que bien après leur fin acoustique effective, alors même que le mot suivant est déjà en cours de traitement. C'est donc l'information acoustique ultérieure qui permet, pour des mots à PU tardif présentés dans un signal de parole continue, d'accéder à l'identification du mot précédent. Dans le cas des mots courts, il est donc la plupart du temps impossible de prédire à l'avance la localisation de la frontière lexicale, même s'ils sont en moyenne plus fréquents que des mots longs. Tant que le système est capable de localiser une frontière lexicale avant de commencer à traiter le mot suivant, le processus est efficace. Cependant, du fait de la quantité importante de mots à PU tardif dans la langue, il est peu probable que cette procédure puisse être fiable dans une situation naturelle de reconnaissance des mots en parole continue. Il faut toutefois noter l'intérêt d'un modèle comme COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987) qui, malgré les difficultés liées à un alignement sélectif sur les débuts de mots, fournit une représentation des processus d'accès au lexique spécifique de la modalité auditive en prenant en considération l'aspect sériel des processus de reconnaissance des mots parlés. Or plusieurs résultats expérimentaux montrent qu'il est important de tenir compte de cette caractéristique du signal de parole (Mattys, 1997).

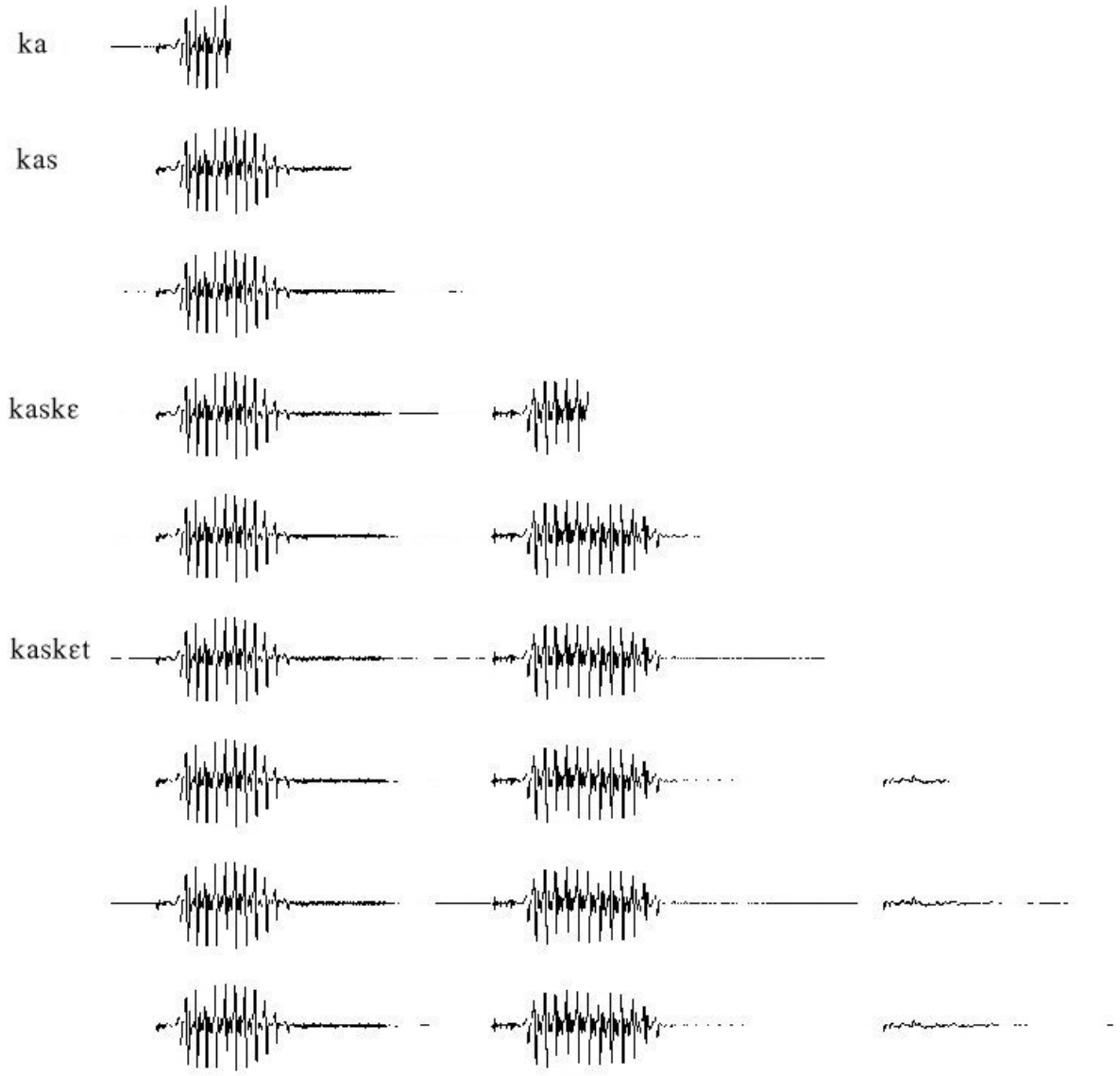

Figure 6 : Exemples de stimuli utilisés dans une tâche de dévoilement progressif (ou *gating*). Ici, le mot utilisé pour créer une série est ‘casquette’. La colonne de gauche indique approximativement les changements de phonèmes perçus.

1.2. Trace

Le modèle TRACE (McClelland & Elman, 1986) s’inspire en grande partie des propositions introduites dans COHORT (et notamment Cohort II, Marslen-Wilson, 1987). Il reprend le principe de sélection progressive des candidats lexicaux à partir de l’accumulation des informations fournies en entrée et implémente le processus de segmentation lexicale comme une conséquence des processus d’identification des mots. La notion de niveau d’activation telle qu’elle est proposée dans COHORT II (Marslen-Wilson, 1987) est centrale dans TRACE (McClelland & Elman, 1986). Il fournit cependant une solution au problème de la segmentation lexicale d’un

signal de parole constitué de mots à Point d'Unicité tardif en introduisant les notions d'*alignement multiple* et de *compétitions lexicales*.

1.2.1. Alignement exhaustif des activations lexicales

Contrairement au principe d'alignement initial des hypothèses lexicales implémenté dans les deux versions de COHORT (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987), McClelland & Elman (1986) introduisent une procédure d'*alignement multiple* (ou *exhaustif*). Dans le modèle COHORT, les hypothèses lexicales (la cohorte de candidats) sont générées uniquement à partir des information acoustiques qui sont considérées comme des débuts de mots. Si le début du signal acoustique correspond au phonème /v/, les éléments lexicaux activés sont ceux qui commencent par /v/ dans le lexique. La séquence de phonèmes /vag/ donnera lieu, dans une première phase, à l'activation de mots comme 'varan', 'vigne', … qui commencent tous par le phonème /v/ (ce qui correspond à 864 mots dans la base de données lexicale Brulex, Content, Mousty, & Radeau, 1990). Un phonème devant nécessairement être considéré comme un début de mot pour générer l'activation d'une cohorte, aucun autre ensemble de candidats ne sera généré tant que le mot commençant par /v/ n'aura pas été identifié, permettant ainsi de déterminer le début du mot suivant et, par conséquent, le phonème qui devra donner lieu à la prochaine cohorte de candidats.

Dans TRACE, au contraire, les hypothèses lexicales peuvent être générées à partir de n'importe quel point du signal. Il n'est pas nécessaire de considérer une partie du signal comme un début de mot pour engendrer une cohorte de candidats. Ainsi, dans le mot /vag/ ('vague'), chaque portion temporelle du signal -de manière simplifiée, chaque phonème[12]- donnera lieu à l'activation d'un ensemble d'hypothèses lexicales contenant ce son. Chaque phonème va donc donner lieu à un ensemble d'hypothèses lexicales dont l'activation ne dépend pas de la localisation préalable d'une frontière entre mots. Ici, une quantité de mots bien plus importante sera activée en raison de cet alignement exhaustif (dans la base de données Brulex, Content et al., 1990, 864 mots commencent par /v/, 2017 par /a/ et 844 par /g/). L'activation des candidats indépendamment d'hypothèses sur les débuts de mots permet de résoudre le problème des mots à PU tardif en faisant reposer la segmentation lexicale sur des phénomènes de compétition entre les candidats activés.

[12] En réalité, et afin de simuler les effets de la coarticulation, le modèle génère un ensemble de candidats tous les 3 cycles de traitement ; ce qui donne donc lieu à plusieurs réseaux lexicaux pour chaque phonème.

1.2.2. Compétitions entre candidats lexicaux

L'utilité d'un alignement exhaustif dans la génération des hypothèses lexicales est intimement liée à la mise en œuvre de phénomènes de compétition entre les candidats lexicaux activés. On entendra ici le mot 'compétitions' dans son acception connexionniste, c'est à dire en termes de régulation du niveau d'activation des différents candidats lexicaux entre eux. Il y a, dans COHORT II (Marslen-Wilson, 1987), un phénomène similaire à ce que l'on peut appeler 'compétitions' : à partir d'une cohorte de candidats lexicaux, le système fait partiellement reposer la sélection finale du candidat sur la fréquence relative des mots qui sont maintenus dans la cohorte, mais à aucun moment les unités lexicales activées ne sont censées modifier le niveau d'activation des autres unités. Cette procédure de pondération des activations en fonction de la fréquence des mots correspond en fait à des processus intégratifs ou décisionnels mais pas à des phénomènes de compétitions lexicales tels qu'ils sont envisagés dans le courant connexionniste. On parlera ici de *sélection lexicale.* La sélection lexicale ne trouve sa source que dans les informations fournies par le flux ascendant d'informations et éventuellement dans les processus de modulation du niveau d'activation des candidats à partir de leur fréquence relative. On réserve l'appellation de *compétitions lexicales* à des modèles dans lesquels les unités lexicales s'influencent mutuellement. Cette influence mutuelle se manifeste par des processus d'interaction entre les représentations lexicales activées (par exemple, inhibitions latérales entre unités lexicales). Ces deux composantes du modèle TRACE (alignement exhaustif et compétitions lexicales) sont désignées comme des processus d'activation et de compétition interactives (IAC, *Interactive Activation and Competition*) qui permettent la propagation des informations à travers les diverses unités (traits acoustiques, phonèmes, mots). Cette propagation de l'activation permet une interaction globale entre les unités et aboutit après un certain délai à une stabilisation de l'état du réseau.

Afin de simuler le décours temporel du traitement de l'information, on laisse s'écouler entre chaque entrée un certain nombre de cycles de traitement et, pour être en mesure de prendre en compte les aspects sériels de la succession des phonèmes dans le signal, on réplique le réseau après un nombre arbitraire de cycles. Après chaque cycle, le niveau d'activation des unités du réseau est enregistré. Une fois qu'un état relativement stable est atteint, on est en mesure d'estimer le comportement du réseau et de savoir quels mots (quels phonèmes, quels traits acoustiques) sont les plus activés à un moment donné du traitement et, par conséquent, quelles sont les frontières lexicales prédites.

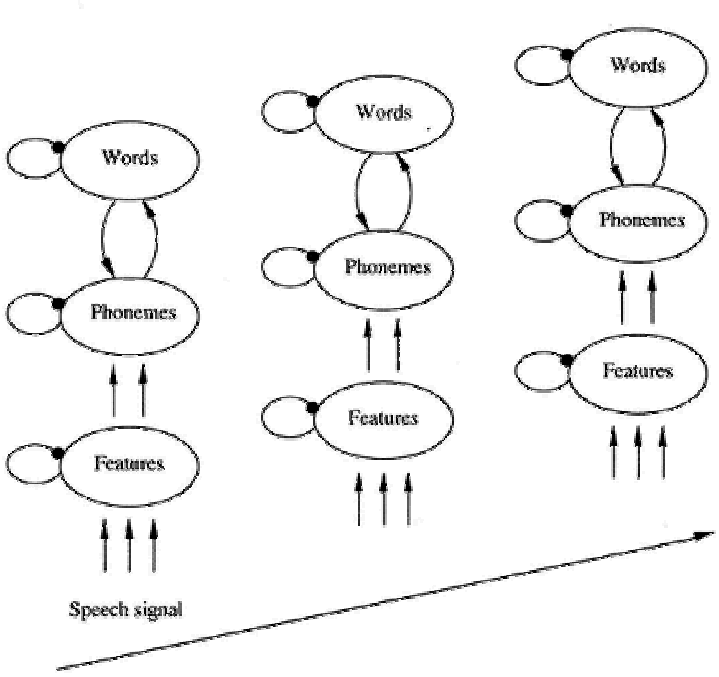

Figure 7 : Représentation graphique du réseau implémenté dans TRACE (McClelland & Elman, 1986). Extrait de Frauenfelder (1996).

1.2.3. Réalisme des procédures implémentées dans TRACE

Le modèle TRACE permet donc de résoudre les difficultés posées par les mots à Point d'Unicité tardif puisque leur reconnaissance ne repose pas sur l'activation d'une cohorte de candidats alignés avec le début de la séquence acoustique mais sur des compétitions entre candidats activés selon une procédure d'alignement exhaustif. Certaines contraintes incitent cependant à poser la question de l'efficacité de ces processus pour l'identification d'une chaîne de mots à partir d'une séquence de phonèmes. L'implémentation des méthodes de représentation sérielle du signal de parole semble par exemple peu adaptée au traitement de séquences de parole relativement longues si le modèle devait utiliser un lexique de taille réaliste. En effet, pour représenter l'ordre des segments dans la séquence de phonèmes, chaque début possible de mot correspond à une réplication du réseau précédent. Les divers réseaux générés sont reliés entre eux afin d'aboutir finalement à une représentation de la distribution des activations en fonction du temps. Plus la séquence à traiter est longue, plus le nombre de réseaux générés par TRACE (McClelland & Elman, 1986) est important. Si l'on souhaite utiliser un lexique contenant autant de mots qu'un locuteur adulte moyen (cinquante à cent mille mots), l'augmentation de la longueur de la séquence à traiter induit un accroissement colossal de la complexité du réseau. Bien que ce paramètre puisse être considéré comme un simple problème algorithmique (cf. la typologie de Marr, 1982) qui ne concernerait que la méthode informatique utilisée pour refléter les capacités du modèle à coder des événements sériels, il serait intéressant d'être en mesure de

traiter ces aspects du signal de parole avec un algorithme plus économique, ne serait-ce que pour alléger la procédure de simulation des données comportementales. Un modèle comme TRACE (McClelland & Elman, 1986) pose cependant des problèmes qui ne se réduisent pas uniquement à un 'manque d'élégance' algorithmique mais qui reposent aussi sur ses aspects computationnels -c'est à dire la théorie du fonctionnement supposé être effectivement mis en œuvre par le système simulé. On remarquera notamment la quantité importante d'homophones dans les langues (surtout en français). Dans la représentation phonémique de la phrase /nuzavjɔ̃kaʃtelɑ̃vlɔpsyrləʃɑ̃/ ('Nous avions cacheté l'enveloppe sur le champ'), El-Bèze (1996) trouve, dans un réseau d'homophones, plus de 3 millions de 'chemins' possibles allant chacun du début à la fin de la phrase[13]. Il est donc coûteux -en termes de puissance de calcul- de retrouver, dans une séquence de phonèmes de longueur moyenne, la suite des mots qui sont effectivement présents. L'obstacle majeur de ce type d'approche est donc en partie lié à la puissance de calcul nécessaire à implémenter cette stratégie dans un système artificiel de perception de la parole. En effet, pour aboutir avec certitude à la séquence de mots qui correspond à l'entrée acoustique, il faut garder une trace mnésique de cette entrée pendant une durée assez longue. Evidemment, il est difficile de comparer la puissance de calcul d'un ordinateur et celle d'un cerveau dans le choix d'un modèle adéquat du fonctionnement cognitif. La question n'est pas de parvenir, dans le domaine de la recherche fondamentale, à développer des modèles informatiques qui seraient en mesure de traiter le signal aussi rapidement que le système cognitif humain avec des ordinateurs qui fonctionnent rarement avec plusieurs processeurs en parallèle. Les critères à adopter doivent au contraire nous conduire à développer des modèles qui reflètent les processus effectivement mis en œuvre par ce système. On peut néanmoins supposer que le système humain a développé des processus cognitifs économiques qui lui permettent de réaliser les tâches avec le moins de complexité possible. On connaît notamment les limites du système de la mémoire de travail. Or la mise en œuvre d'un système de segmentation lexicale qui reposerait uniquement sur des processus de compétitions nécessiterait le recours à une mise en mémoire parfois vertigineuse d'une partie des traitements qui ont précédé. Ceci induit à remettre en question la validité d'une approche consistant à faire reposer presque exclusivement la localisation des frontières de mots sur les processus d'accès au lexique.

Il est cependant clair que les processus de compétition lexicale sont partie intégrante du système d'accès au lexique. Plusieurs travaux mettent ainsi en évidence des effets du nombre de

[13] Sans prendre en compte la plausibilité syntaxique ou sémantique de la séquence qui en découle. Ce nombre peut sembler énorme. Il faut cependant noter qu'il prend en compte les variantes de prononciation. On peut penser que, même sans ces variantes, on aboutirait à une quantité considérable de 'chemins' possibles.

compétiteurs lexicaux ou de l'existence de compétiteurs plus fréquents dans les latences de reconnaissance des mots (McQueen, Norris, & Cutler, 1994; Norris, McQueen, & Cutler, 1995). Il pourrait être intéressant de diminuer la puissance de calcul nécessaire à la mise en compétition d'un nombre aussi considérable d'éléments du lexique sans pour autant supprimer cette particularité du système.

1.3. Shortlist

Pour aboutir à une segmentation lexicale, le modèle SHORTLIST (Norris, 1994) offre des procédures de traitement à la fois plus économiques et réalistes que celles implémentées par McClelland & Elman (1986) dans TRACE. Cette économie dans la puissance de calcul repose sur deux différences : le recours à des boucles de récurrence (Jordan, 1986; Elman, 1990 ; cf. aussi Chapitre 1, Section 2.1.2.2) et la restriction du nombre de candidats activés. Norris, McQueen, & Cutler (1995) ont également implémenté des procédures prélexicales fondées sur une classe de régularités phonologiques -les alternances accentuelles- afin d'influencer les activations lexicales et d'accélérer le processus de sélection des candidats appropriés.

1.3.1. Représentation de l'ordre par récurrence

Au moment de la publication de l'article qui présentait le modèle TRACE (McClelland & Elman, 1986), il n'existait pas encore de procédure économique pour traiter les aspects temporels et / ou sériels de l'entrée d'un modèle. Nous avons vu dans la partie consacrée à TRACE qu'une réplication des réseaux d'activation avait été mise en œuvre pour coder l'ordre des segments phonémiques et leur position dans les mots. Cette méthode induit une croissance exponentielle de la charge de traitement lorsqu'on augmente simultanément la taille de la séquence et celle du lexique. La même année, Jordan (1986) publie un rapport présentant la notion de récurrence locale dans un réseau de neurones. Cette procédure consiste à renvoyer aux unités d'entrée l'information reçue au cours du cycle de traitement précédent par le biais d'une couche d'unités supplémentaire. L'état du système au temps `t` est alors en partie déterminé par son état au temps `t-1`. De fait, lorsque les unités intermédiaires reçoivent l'information fournie par les unités d'entrée, elles disposent également d'une 'mémoire' du cycle de traitement qui a précédé (cf. Figure 8). La méthode de représentation de la structure temporelle d'une séquence par le biais du principe de récurrence est présentée dans Elman (1990). Cette procédure peut être implémentée dans un Réseau Ascendant (*Feedforward Network*) qui prend alors le nom de Réseau Récurrent Simple (*Simple Recurrent Network* -SRN-, cf. MERGE, Chapitre 1). Cette méthode de récurrence de l'information avec insertion d'un délai fournit une mémoire temporelle au modèle qui lui

permet de coder à la fois les durées et l'ordre des phonèmes ainsi que leur position dans les mots en activant un seul réseau lexical. Dans SHORTLIST, cette procédure est *simulée* par une recherche classique dans un dictionnaire afin de générer un ensemble d'hypothèses lexicales unique pour l'ensemble des points du signal[14]. Les SRN présentent cependant des propriétés spécifiques qui n'en font pas des candidats adéquats pour simuler les processus de reconnaissance des mots.

1.3.2. Réseaux récurrents simples et enchâssement lexical

Si un réseau récurrent simple est entraîné à reconnaître 'car' et 'cargo' au cours de la phase d'apprentissage, la présentation de la séquence /kargo/ en test lui permettra dans une première étape de reconnaître 'car'. Lorsque /go/ est présenté, le niveau d'activation de l'unité correspondant au mot 'cargo' augmente mais le SRN n'est pas conçu pour réduire l'activation de 'car'. Ce type de modèle est uniquement capable de donner une 'interprétation' de la séquence au moment où il reçoit l'entrée. Il ne peut pas revenir en arrière pour réduire le niveau d'activation de 'car' à partir de l'information acoustique subséquente dans le but de sélectionner l'interprétation la mieux adaptée. Si l'on imagine que les niveaux d'activation reflètent la réponse qui serait donnée par un auditeur, le modèle *identifie* d'abord 'car' puis, avec le traitement du contexte subséquent, 'cargo' ; mais à aucun moment il ne considère 'car' comme faisant partie intégrante du mot 'cargo'. Malgré la possibilité qu'ont les mots enchâssés d'être

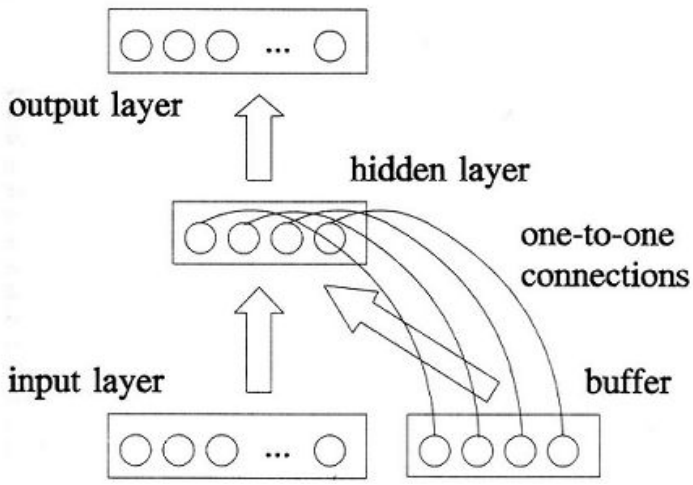

Figure 8 : Illustration d'une boucle de récurrence locale dans un réseau neuronal (extrait de Murre & Goebel, 1996).

[14] On voit ici que l'enjeu n'est pas tant de fournir des procédures économiques de traitement (le codage de l'ordre dans un SRN est plus économique que celui implémenté dans TRACE mais la procédure d'apprentissage pourrait durer plusieurs semaines avec un lexique de taille moyenne) que d'implémenter une représentation réaliste du codage temporel et sériel des événements dans un modèle du système cognitif.

transitoirement activés (Isel & Bacri, 1999 ; Vroomen & De Gelder, 1997), le système de traitement de la parole aboutit toujours à un choix pour le découpage lexical et l'identification des mots présents dans l'énoncé. Avec un SRN, tout se passe comme si nous percevions dans la chaîne parlée toutes les combinaisons possibles de mots sans être en mesure de choisir la bonne. Ces réseaux constituent une classe de modèles bien adaptée à l'identification phonémique car un phonème ne peut pas en contenir un autre. Dans le cas de la reconnaissance des mots, ils doivent inclure des procédures spécifiques qui les conduisent à considérer chaque partie de l'entrée comme provenant d'un mot unique. Dans SHORTLIST, le résultat de ce qui devrait être simulé par un SRN (mais qui repose en fait sur une recherche classique dans un lexique informatisé) est transféré vers un réseau d'Activation et de Compétition Interactive similaire à TRACE. Le nombre de candidats lexicaux qui sont envoyés au réseau IAC est cependant restreint par une limitation arbitraire du nombre de compétiteurs (c'est la shortlist, qui est composée, dans les simulations présentées par Norris, 1994, de 3 à 30 candidats). Ils entrent alors en compétition les uns avec les autres dans un réseau lexical identique au niveau lexical que l'on trouve dans TRACE de manière à aboutir à l'identification d'une unique séquence de mots. Cette restriction de la taille de l'ensemble des hypothèses lexicales et du nombre de réseaux générés en parallèle -qui se réduit à un réseau unique- permet d'utiliser un lexique bien plus grand (environ 50 000 mots) dans les simulations et de conserver la mise en œuvre des phénomènes de compétition entre candidats lexicaux comme une caractéristique essentielle des processus de segmentation de la parole en mots.

1.3.3. Implémentation de procédures de segmentation prélexicales

Plus récemment, des procédures de segmentation prélexicales ont été implémentées dans ce modèle afin de guider la reconnaissance des mots à partir de règles dérivées des régularités phonologiques de l'anglais. En effet, des données comportementales ont mis en évidence le recours à des informations prélexicales pour segmenter le signal de parole en mots. Des connaissances sur certaines régularités des langues (phonologiques ou distributionnelles) permettraient de guider les processus de segmentation de la parole et pourraient faciliter la tâche du système. Ces effets sont simulés par une modification des activations lexicales des divers candidats en fonction de leur alignement avec des *formes phonologiques* typiques de la langue Une description plus approfondie de cette catégorie de processus est présentée dans la section suivante.

2. Indices de segmentation prélexicaux

Pour se libérer des contraintes liées à des processus de segmentation lexicale qui ne reposeraient que sur des procédures d'accès au lexique, ainsi que pour mieux rendre compte des données comportementales ou acoustiques, un certain nombre de propositions ont été développées dans lesquelles il est envisagé de recourir à des informations prélexicales pour guider la segmentation d'un signal de parole en mots. On trouve, parmi ces propositions, une variété d'indices qui permettraient de localiser des frontières de mots avec plus ou moins de certitude. L'approche la plus courante a consisté à proposer des indices signalant des frontières. Il a également été envisagé, assez rarement il est vrai, de recourir à des indices indiquant la relative nécessité de regrouper certaines séquences à l'intérieur d'une hypothèse lexicale. Ces deux approches fournissent cependant un intérêt équivalent dans la description de modèles de la segmentation lexicale : la première se fixe pour objectif de séparer des séquences en localisant des frontières (ségrégation) alors que la seconde cherche des indices impliquant la nécessité de regrouper certaines portions du signal pour la recherche lexicale (intégration). Quoiqu'il en soit, on peut classer ces indices dans deux catégories : des informations segmentales qui sont fournies par les caractéristiques acoustico-phonétiques ou phonologiques individuelles des sons ; et des informations suprasegmentales qui impliquent la mise en relation d'indices acoustico-phonétiques ou phonologiques portés par plusieurs segments ou groupes de segments de la chaîne parlée. Dans tous les cas, les indices proposés présentent une certaine régularité qui permettrait de les utiliser comme des prédicateurs de la présence ou de l'absence d'une frontière de mots.

2.1. Indices segmentaux

Parmi les indices segmentaux, dont la particularité est de fournir des informations utilisables sans qu'il soit nécessaire de les mettre en rapport avec d'autres éléments du signal, les indices majeurs de frontières lexicales sont les variations allophoniques. Les allophones sont des variantes acoustiques d'un même phonème. Ils n'ont pas de valeur distinctive dans la langue mais présentent des variations acoustiques qui peuvent, dans certaines conditions, adopter un comportement stable. On observe notamment des régularités en fonction de la structure de la syllabe dans laquelle est prononcé le phonème (en français, /o/ s'ouvre et aboutit au phonème [ɔ] dans une syllabe fermée). La prononciation peut aussi dépendre de la position du phonème dans la syllabe (en français, /r/ se prononce différemment en attaque ou en coda). Leur prononciation

varie donc en fonction de paramètres précis sans induire de distinctivité phonémique. Ces variantes acoustiques constituent des représentants de la même catégorie phonémique.

2.1.1. En anglais

Les indices qui ont été mis en évidence dans des tâches perceptives se rapportent à la langue anglaise. Nous verrons dans la section suivante que des indices similaires pourraient intervenir en français. Les travaux présentés montrent que certaines régularités allophoniques peuvent être utilisées par des locuteurs de la langue comme indicateurs de frontières de mots. En anglais, il existe par exemple une tendance à prononcer les phonèmes en position initiale de mots avec des caractéristiques particulières qui les distinguent des mêmes phonèmes prononcés en position finale ou médiane. Ces différences de prononciation peuvent être utilisées par des locuteurs pour identifier une séquence de parole ambiguë (Nakatani & Dukes, 1977). Les auteurs montrent, dans une analyse acoustique de séquences lexicalement ambiguës, que le signal de parole est porteur d'indices différenciant les phonèmes selon qu'ils sont prononcés en début de mot ou dans d'autres positions. Ainsi, les consonnes occlusives sont glottalisées ou laryngealisées en position initiale : le /n/ médian de 'no n<u>o</u>tion' -qui se trouve à l'initiale du second mot- a des caractéristiques acoustiques différentes du /n/ médian de 'know<u>n</u> ocean' -qui est localisé à la fin du premier mot. Cette modification des caractéristiques acoustiques des consonnes en fonction de leur position lexicale permet à des auditeurs d'identifier correctement la séquence effectivement produite malgré son ambiguïté phonologique (avec 70 à 90% de taux de correspondance entre la séquence lexicale effectivement produite et la réponse des auditeurs). Par ailleurs, c'est la laryngealisation ou la glottalisation de la consonne en début de mot qui constitue un indice pertinent ; l'absence de cette caractéristique n'induit pas nécessairement que le phonème soit en position médiane ou finale de mot.

Outre les variations allophoniques dépendantes des frontières lexicales, il existe aussi des modifications acoustiques des phonèmes en fonction de leur position dans la syllabe. Church (1987) propose d'intégrer des connaissances sur les régularités des alternances allophoniques liées à la structure syllabique dans les processus de 'segmentation' (plus précisément de *parsing*) lexicale. Ce type d'informations semble effectivement être mis en œuvre par le système de traitement de la parole. Ainsi, Nakatani & Dukes (1977) observent un rôle de la régularité des variations allophoniques du /r/ et du /l/ anglais. Ces deux phonèmes obéissent à des contraintes de prononciation différentes selon qu'ils sont prononcés en attaque ou en coda syllabique et leurs données comportementales montrent que les locuteurs anglophones se fondent sur ce type d'informations pour identifier correctement les séquences ambiguës. Il existe cependant une

relation, dans l'échantillon des séquences qu'ils ont analysées, entre les découpages syllabique et lexical. La distribution des allophones de ces deux phonèmes est déterminée par la position dans les mots mais, contrairement au phénomène précédent, cette catégorie de variations allophoniques peut en réalité indiquer aussi bien un début qu'une fin de mot puisque découpage syllabique et lexical ne correspondent pas nécessairement. Ce type d'information est donc moins fiable que la laryngealisation des occlusives en début de mot puisqu'il repose sur des contraintes syllabiques. Ces régularités allophoniques ne sont entièrement fiables que si les découpages syllabique et lexical correspondent. Il reste cependant que ce type d'informations peut servir de pondérateur dans les choix effectués par le système de segmentation lexicale et le modèle proposé par Church (1987) implémente un recours à ce type d'informations pour faciliter les processus de reconnaissance des mots.

Il conviendra toutefois de rester prudent quant à l'interprétation que l'on pourrait être tenté de donner de ces données expérimentales. En effet, Nakatani & Dukes (1977) ont bien montré que, placés dans une situation de choix d'une signification lorsqu'ils écoutent des séquences phonologiquement ambiguës, des informations allophoniques peuvent être prises en compte par le système pour faire le choix correct. Ce résultat ne permet cependant pas d'en conclure un quelconque rôle des indices allophoniques dans les processus mis en place en temps réel pour segmenter un signal de parole continue en mots. En effet, lorsque le système cognitif traite un signal de parole dans une situation de communication, les délais dont il dispose pour analyser et comprendre la suite de mots sont très restreints. Les temps de traitement inhérents à ces situations sont extrêmement courts. Or, dans l'expérience de Nakatani & Duke (1977), les participants disposaient de délais de réponse relativement longs et pouvaient donc potentiellement utiliser des indices qu'ils n'auraient pas été en mesure de prendre en compte dans une situation plus contrainte en termes de délais de traitement.

2.1.2. En français

Il existe en français des régularités similaires à celles de la langue anglaise qui incitent à penser que des processus semblables pourraient prendre place dans le système de segmentation de la parole chez les locuteurs de cette langue. Certains phonèmes du français ont par exemple des variantes allophoniques dont les caractéristiques dépendent en partie de leur position dans la syllabe. Le phonème /r/ est la fricative uvulaire voisée [ʁ] en attaque syllabique et la fricative uvulaire non-voisée [χ] en coda. De même qu'en anglais, la position syllabique ne permet pas de prédire avec certitude la position lexicale d'un phonème. Si, dans une séquence constituée de deux mots, le premier se termine par une consonne et le second commence par une voyelle, la

consonne finale du premier mot est enchaînée avec la voyelle initiale du second mot et peut subir les règles de syllabation propres à la langue indépendamment du découpage lexical. La position lexicale du phonème ne peut donc pas être totalement déterminée par ce phénomène d'allophonie du /r/ dans le cas d'un signal de parole continue. Celui-ci permettrait cependant, lorsque le système de reconnaissance de la parole rencontre un /ʁ/ voisé de savoir qu'il est en début de syllabe et, par conséquent, qu'il est possible qu'il soit un début de mot ; l'occurrence d'un /χ/ non-voisé indiquant au contraire une fin de syllabe et, potentiellement, une fin de mot. Cette information pourrait être prise en compte, parmi d'autres, pour générer des hypothèses sur le découpage lexical sans que les décisions ultimes reposent uniquement sur cet indice. A notre connaissance, aucune étude n'a cependant été réalisée afin d'étudier avec précision le rôle des indices allophoniques dans les processus de segmentation de la parole en français.

2.2. Indices suprasegmentaux

La majorité des indices qui ont été proposés comme marqueurs potentiels des frontières lexicales (en anglais aussi bien qu'en français ou en néerlandais) sont en fait de nature suprasegmentale. Cette caractéristique implique la mise en relation des informations portées par plusieurs segments ou groupes de segments de la chaîne parlée. La comparaison des indices entre eux, ou les caractéristiques spécifiques d'une *séquence* de segments, fournit une information sur l'éventuelle présence ou absence d'une frontière lexicale. Nous décrirons ici 3 types d'indices suprasegmentaux : métriques, phonotactiques et distributionnels.

2.2.1. Indices métriques

Les indices métriques sont portés par les variations prosodiques du signal de parole. Celles-ci se manifestent par des changements d'intensité, de durée et de hauteur de la fréquence fondamentale (F_0) qui présentent une régularité dans leur mode de succession. Ces alternances métriques dépendent essentiellement de deux paramètres : les informations prosodiques attachées aux représentations lexicales qui déterminent l'accentuation à adopter lorsqu'on prononce un mot et les contraintes prosodiques liées à la structure phonologique de la langue qui ont pour domaine d'application ce qu'on appelle le groupe prosodique -groupe qui n'est pas nécessairement équivalent au mot. Ces deux composantes de la prosodie constituent deux types d'indices suprasegmentaux qui pourraient jouer un rôle dans les processus de segmentation lexicale.

2.2.1.1. Accent lexical

La langue anglaise se caractérise par l'existence de paramètres accentuels liés aux représentations lexicales (Fear, Cutler, & Butterfield, 1995). Dans les mots anglais, on observe une alternance de syllabes accentuées (*strong syllables*, syllabes fortes) et non accentuées (*weak syllables*, syllabes faibles). Cette alternance étant déterminée par les représentations lexicales, elle a un caractère distinctif : la position de l'accent dans une séquence de phonèmes peut déterminer sa signification. Cet accent lexical présente cependant une régularité importante en ce qui concerne la position qu'il occupe dans les mots. Cutler & Carter (1987) montrent ainsi qu'en anglais, 80% des mots d'un lexique informatisé portent l'accent sur leur syllabe initiale. Dans un système de segmentation de la parole en mots qui chercherait des informations acoustiques aptes à indiquer des frontières de mots, la prise en considération de cette régularité permettrait de privilégier les syllabes fortes comme des débuts de mots.

C'est effectivement le type de procédures qui semblent être mises en œuvre par le système de traitement de la parole chez des locuteurs anglais. Cutler & Norris (1988) ont pu mettre en évidence un rôle des syllabes fortes dans les processus de segmentation lexicale avec une tâche de *word-spotting* (extraction de mot). Cette tâche consiste à présenter à des auditeurs des séquences de parole sans signification (des non-mots) dans desquelles il est parfois possible d'isoler une suite de phonèmes désignant un mot dans la langue. La suite /mɑ̃tɛʃ/ par exemple n'a aucun sens en français ; mais si l'on cherche un mot de la langue dans cette séquence, on peut isoler le mot 'menthe' en position initiale. Cutler & Norris (1988) ont introduit ce nouveau paradigme expérimental avec une expérience dans laquelle ils demandaient à des locuteurs anglais de détecter des mots monosyllabiques dans des non-mots bisyllabiques. Les mots étaient tous des monosyllabes accentués (comme la plupart des monosyllabes de classe ouverte, par exemple *mint*, 'menthe') et apparaissaient en position initiale du non-mot. Les caractéristiques accentuelles de la séquence VC qui suivait variaient afin de tester la différence entre l'occurrence d'une syllabe forte (/ɘɪv/ dans /mintɘɪv/) et celle d'une syllabe faible (/ɘʃ/ dans /mintɘʃ/). Les sujets devaient appuyer le plus rapidement possible sur un bouton réponse lorsqu'ils détectaient un mot de la langue dans les non-mots qui leur étaient présentés. Les mots qui apparaissaient au début des non-mots n'étaient pas préalablement indiqués aux participants[15]. Cutler & Norris (1988) montrent que les locuteurs anglais éprouvent plus de difficulté à reconnaître le mot initial lorsque la seconde syllabe est forte (/mintɘɪv/) que lorsque celle-ci est

[15] C'est la raison pour laquelle on traduit *word-spotting* par 'extraction' plutôt que par 'détection' de mot.

faible (/mintəʃ/). Il est donc plus difficile d'identifier un mot qui porte l'accent lexical lorsque celui-ci est prononcé dans une séquence *strong-strong* (SS) que dans une suite *strong-weak* (SW). Selon les auteurs, cet effet serait lié à un alignement des activations lexicales sur le début des syllabes fortes. Dans les séquences *strong-strong*, un ensemble de candidats serait activé à partir du /m/, puis un autre à partir du /t/. Du fait de l'alignement d'un second réseaux de candidats sur le /t/ et du recouvrement des deux réseaux, des phénomènes de compétitions lexicales (cf. Section 1.2.2) devraient s'opérer entre les membres de ces réseaux pour aboutir à l'identification du mot *mint*. Ce recouvrement des activations conduirait à éprouver des difficultés pour rattacher le /t/ aux trois premiers phonèmes qui le précèdent (/min/). Au contraire, lorsque la seconde syllabe est faible, seul le phonème initial serait utilisé pour générer les activations et on n'observerait alors aucun conflit entre deux réseaux de candidats partageant un même phonème. Des résultats similaires ont été obtenus avec des locuteurs néerlandais (Vroomen & de Gelder, 1995) effectuant la même tâche.

2.2.1.2. Alternances prosodiques

L'accent lexical ne constitue cependant pas la seule composante prosodique dans le cadre de l'étude des phénomènes intonatifs. Outre les possibilités de stockage lexical des informations prosodiques qui se manifestent en anglais, la chaîne parlée subit des contraintes prosodiques qui sont indépendantes du lexique. Les théories prosodiques développées en phonologie proposent une représentation de la chaîne de parole en termes de composants prosodiques hiérarchisés (Nespor & Vogel, 1983 ; Rossi, 1999 ; Vaissière, 1983) dont le domaine d'application peut s'étendre au-delà des mots individuels. Une formulation typique de la hiérarchie prosodique consiste à regrouper les phonèmes en syllabes, elles-mêmes regroupées en pieds (*feet*). Ces pieds constituent des 'mots phonologiques' (*phonological words*). Ceux-ci ne correspondent pas nécessairement à des 'mots' à l'exemple de ceux que l'on peut trouver dans un dictionnaire mais regroupent le plus souvent des segments provenant de plusieurs unités lexicales. A partir de cette structure prosodique complexe de la phrase, Grosjean & Gee (1987) proposent, dans le même esprit que Cutler & Norris (1988), de faire reposer les processus d'accès lexical en anglais sur les syllabes fortes. Ils envisagent cependant, en raison de l'absence de correspondance biunivoque entre découpages lexical et prosodique, des procédures plus souples dans l'agencement temporel de l'accès aux représentations segmentales intégrées dans les syllabes non-accentuées par le biais de retours en arrière dans les processus d'intégration des syllabes fortes et faibles en mots. Les syllabes accentuées donneraient lieu à un ensemble d'activations lexicales qui ne se restreindraient pas aux syllabes initiales mais à toute syllabe, quelle que soit sa position dans les

mots. Cette première étape de génération d'un réseau de candidats donnerait lieu à une étape d'identification des syllabes non-accentuées environnantes -qui pourrait reposer sur des procédures de traitement aussi bien ascendantes que descendantes. La charge de travail induite par les processus de reconnaissance des mots serait restreinte en raison de la limitation des réseaux d'activation aux mots qui contiennent cette syllabe accentuée. Les mots effectivement présents dans l'énoncé seraient alors identifiés par le biais d'une combinaison des informations dérivées des réseaux d'activation et de l'identification des syllabes non-accentuées.

Ce type de procédures pourrait non seulement s'appliquer à l'anglais mais également -en raison de l'universalité des variations intonatives dans les langues- à toute langue du monde, que l'accent ait ou pas de valeur distinctive. En français par exemple, l'accent n'est pas dépendant d'informations stockées dans le lexique mais repose essentiellement sur les contraintes phonologiques mises en œuvre dans la structuration prosodique des énoncés. Son domaine d'application, le groupe prosodique ('syntagme phonologique' ou 'mot phonologique'), le rend dépendant des paramètres intonatifs de la langue sans être lié aux représentations lexicales. Ces variations prosodiques se manifestent par des modifications de hauteur, d'intensité ou de durée relative des syllabes. Dans la phrase, 'Le menuisier a scié une planche et l'a rabotée', on peut distinguer trois groupes prosodiques :

[Le menuisier] [a scié une planche] [et l'a rabotée].

Chacun de ces groupements prosodiques se manifeste par une élévation progressive -suivie d'une diminution- de la F_0 ainsi que par une accentuation de la syllabe finale. Cette accent final se manifeste par un *allongement de la durée* et / ou une *augmentation de l'intensité acoustique* et / ou une *rapide élévation de la* F_0 de la syllabe qui le porte. Du fait de ses caractéristiques, l'accent prosodique français semble être déterminé à remplir une fonction totalement différente de celle qu'exerce l'accent lexical anglais. Tout d'abord, il n'est pas porté par des unités initiales mais finales. En outre, son domaine d'application n'est pas le mot mais le groupe prosodique. On pourrait penser que cette dernière caractéristique le rend difficilement apte à fournir des indices pertinents pour la segmentation d'un signal de parole en mots. Il reste cependant un indice potentiel pour la structuration du signal de parole et la reconnaissance des mots si l'on se réfère au modèle proposé par Grosjean & Gee (1987). En effet, il présente une stabilité beaucoup plus marquée que l'accent anglais qui, bien qu'il soit porté dans la majorité des mots par la syllabe initiale, peut être se porter sur d'autres positions lexicales dans un nombre non-négligeable de mots. L'accent français ne présente pas cette instabilité. Toute phrase prononcée par un locuteur natif subit les contraintes intonatives du français et se trouve découpée en groupes prosodiques

qu'il est possible de déterminer par une analyse des paramètres acoustiques propres à cet accent (intensité, durées, F_0). L'occurrence et la détection de cet accent indiquerait sans ambiguïté la fin d'un groupe prosodique.

Banel & Bacri (1997) ont utilisé la tâche de *word-spotting* introduite par Cutler & Norris (1988) afin de tester le rôle de ces régularités intonatives dans les processus de segmentation lexicale en français. Les auteurs ont fait écouter à des locuteurs natifs des non-mots de deux syllabes dont certains commençaient par un mot monosyllabique (les autres ne contenant aucun mot de la langue). Les stimuli expérimentaux présentaient 3 types de configuration intonative : *ïambique* (le pattern classique du français, le phénomène d'accentuation étant localisé sur la syllabe finale du stimulus), *trochaïque* (l'inverse du pattern ïambique, la première syllabe portant seule l'accent), et neutre (une *spondée* dans laquelle les deux syllabes du stimulus sont accentuées). Le paramètre intonatif manipulé était la *durée* : les séquences ïambiques étant constituées d'une syllabe courte suivie d'une syllabe longue (pattern court-long), les stimuli trochaïques étant de type long-court et les spondées présentant un pattern long-long. Le pattern trochaïque (long-court) donne lieu à des délais d'extraction du mot initial plus courts que le pattern ïambique (court-long). Dans la suite de phonèmes /lɑ̃pzɔk/, le mot 'lampe' est détecté assez rapidement lorsque la syllabe /lɑ̃p/ est plus longue que la syllabe /zɔk/. Les latences de détection du mot 'lampe' sont au contraire plus longues lorsque la syllabe /lɑ̃p/ est plus courte que la syllabe suivante. Or, du fait de la localisation de l'accent sur la dernière syllabe des stimuli, le pattern ïambique respecte les contraintes intonatives du français. Dans la configuration trochaïque, au contraire, les paramètres intonatifs correspondent à une rupture prosodique entre la première et la seconde syllabe des stimuli. Les auteurs en concluent que les régularités prosodiques du français *exprimées dans les durées syllabiques* jouent un rôle dans les processus de segmentation de la parole en français. La routine de traitement mise en œuvre consisterait à regrouper les segments de la chaîne de parole en fonction des durées relatives des syllabes, les syllabes longues indiquant la fin d'un groupe prosodique permettraient de structurer le signal perçu et d'empêcher toute tentative d'identification d'un candidat lexical qui ne respecterait pas le découpage prosodique induit par l'allongement de la syllabe finale.

2.2.2. Indices probabilistes

Les résultats présentés en faveur d'un rôle des indices probabilistes dans la segmentation du signal de parole en mots ont été obtenus, chez les locuteurs adultes aussi bien que chez des nouveau-nés, avec des expériences d'apprentissage de langues artificielles utilisant divers paradigmes expérimentaux adaptés à l'âge des participants. Chez l'enfant, comme chez l'adulte,

on observe une aptitude à traiter les différences de fréquence des séquences de phonèmes qui pourrait être utile dans les processus de localisation des frontières entre mots.

2.2.2.1. Les travaux sur les langues artificielles

L'approche introduite pour étudier le rôle des indices probabilistes dans les processus de segmentation de la parole a consisté à présenter de longues séquences de parole sans signification dans lesquelles on manipule la fréquence d'apparition de certaines suites de phonèmes. Si les probabilités de succession des phonèmes dans le signal sont traitées par le système pour émettre des hypothèses sur les frontières de mots, on s'attend à ce que les participants présentent, dans une tâche ultérieure, des comportements qui reposent sur les différences de probabilité de ces séquences. Des résultats confirmant l'importance des régularités probabilistes ont été mis en évidence chez des nouveau-nés de quelques jours à qui on a fait écouter des séquences de parole sans signification. On observe, dans une procédure de conditionnement induisant les enfants à tourner la tête vers un jouet lorsqu'ils entendent une séquence familière (*head-turning conditioning procedure*), que les stimuli présentant des probabilités transitionnelles élevées donnent lieu -lorsqu'ils sont présentés isolément- à des taux de préférence plus importants que ceux dont les probabilités transitionnelles sont faibles. Des nouveau-nés sont donc en mesure, dès leur premier contact avec le langage, de calculer les fréquences d'apparition de séquences de phonèmes.

Des processus semblables sont observables chez l'adulte. Des locuteurs anglophones ont été exposés pendant quelques minutes à des séquences de parole sans signification dans lesquelles on avait manipulé les probabilités transitionnelles des paires de phonèmes (Saffran, Newport et al., 1996). Dans une phase ultérieure, les participants écoutaient des stimuli extraits des séquences qui leur avaient été présentées. Leur tâche était de donner un 'jugement lexical' consistant à dire s'ils pensaient que le stimulus correspondait ou pas à un mot dans la langue imaginaire qu'ils venaient d'entendre. On observe une tendance à faire reposer les réponses sur les probabilités transitionnelles des suites de deux phonèmes. Les valeurs faibles de probabilités transitionnelles incitent les participants à supposer qu'une frontière de mots sépare les deux phonèmes alors que les valeurs élevées les conduisent à considérer ces suites comme appartenant au même mot. Les auteurs en concluent que des informations de type probabiliste peuvent être utilisées par le système cognitif dans les processus de traitement de la parole. Des propositions similaires se retrouvent chez d'autres auteurs (Brent & Cartwright, 1996; Brent, 1997) qui présentent des simulations mettant en évidence la facilitation des processus de segmentation

lexicale liée à l'utilisation de données probabilistes sur les régularités de co-occurrence des phonèmes.

2.2.2.2. Traitement ou mémorisation ?

La phase préalable de traitement d'une longue séquence de parole sans signification dans les études sur le rôle des indices probabilistes dans les processus de segmentation lexicale pose le problème du locus des effets observés. Cette première phase, qui a pour objet de permettre aux participants de calculer la fréquence des groupes de phonèmes, induit en effet des phénomènes d'apprentissage. Il est nécessaire, pour que des effets probabilistes émergent dans la phase test, de mémoriser les probabilités d'apparition des séquences dans la phase d'acquisition. On n'est donc pas en mesure de savoir si, dans le cadre des processus mis en œuvre par le système cognitif dans ces expériences, on met en évidence un recours aux probabilités dans les processus de mémorisation ou de traitement proprement dit. L'interprétation qui a été donnée de ces résultats est que le système cognitif est en mesure d'utiliser les probabilités transitionnelles des suites de deux phonèmes pour localiser les frontières de mots. Ceci impliquerait que les participants utilisent essentiellement l'information apportée par les séquences rares pour repérer les frontières probables. On peut cependant envisager que les effets de facilitation de la reconnaissance dans la seconde phase reposent en réalité sur une mémorisation plus facile des séquences qui sont présentées fréquemment dans la première phase. Les effets probabilistes observés pourraient alors s'expliquer par une mémorisation plus facile de ce que l'on entend souvent plutôt que par la mise en œuvre de processus consistant à localiser des frontières à partir des informations fréquentielles.

Van der Lugt (1999) a étudié l'effet de la fréquence des paires de phonèmes CV et VC dans les processus de segmentation de la parole en utilisant comme indice de probabilité la fréquence de ces suites dans le lexique néerlandais. Du fait de la comparaison de chaînes phonémiques de fréquences diverses dans la langue, il n'est pas nécessaire d'avoir recours à une phase préalable d'apprentissage afin de faire éventuellement émerger des effets probabilistes. C'est une tâche de *word-spotting* qui a été utilisée dans cette expérience. Les participants, des locuteurs néerlandais, devaient identifier des mots d'une syllabe en position initiale de non-mots bisyllabiques. La fréquence des séquences initiales CV et finales VC du mot était manipulée. On s'attend à ce que, si les probabilités d'apparition des paires de phonèmes sont effectivement utilisées par le système de segmentation de la parole, il soit plus facile d'isoler un mot contenant une séquence fréquente qu'un autre contenant une séquence rare. Les données obtenues ne permettent de conclure qu'à un effet limité de la fréquence. Lorsque la fréquence est manipulée

dans la syllabe finale (qui n'appartient pas au mot à extraire) ou en position finale du mot (suite VC), on n'observe aucun effet de la fréquence. Un effet de la fréquence émerge cependant en position initiale du mot. C'est donc la fréquence de la séquence initiale CV qui influence les latences d'extraction de mot. Il semble ainsi que les effets probabilistes observés dans les expériences de langue artificielle ne puisse s'expliquer totalement par un rôle de la fréquence dans la segmentation mais soient en réalité interprétables en termes mnémoniques. On remarquera cependant que les chaînes manipulées dans les expériences de Van Der Lugt (1999) présentent des caractéristiques particulières : elles ne se situent jamais à la frontière lexicale. Ainsi, les séquences dont la fréquence est étudiée sont intégrées soit à la partie qui correspond au non-mot, soit à celle qui correspond au mot. Jamais la fréquence des séquences qui chevaucheraient la partie correspondant au mot et le reste du stimulus n'est manipulée alors que c'est spécifiquement la probabilité d'occurrence de cette séquence qui pourrait être le plus utile au déclenchement d'une segmentation fondée sur les probabilités.

2.2.3. Indices phonotactiques

Outre les contraintes prosodiques -qui gouvernent les paramètres intonatifs dans la production de la parole- et les régularités probabilistes -qui concernent la fréquence d'apparition d'une chaîne de phonèmes dans la langue-, les langues obéissent à un certain nombre de contraintes dites phonotactiques. Cette classe de contraintes détermine pour chaque langue l'agencement séquentiel des phonèmes dans les syllabes, les morphèmes et les mots. On considérera que les régularités probabilistes dont nous avons parlé dans la section précédente sont distinctes des contraintes phonotactiques dont il est question ici. Rapidement, les régularités probabilistes nous semblent liées à des fréquences d'apparition quelle que soit leur position dans les unités phonologiques classiques (syllabes, morphèmes, mots). Les régularités phonotactiques impliquent au contraire une référence à la position qu'elles occupent dans ces mêmes unités. Une discussion plus approfondie de la notion de contraintes phonotactiques sera développée dans le Chapitre 3 afin de justifier cette distinction. Dans cette section, nous décrivons un certain nombre de données expérimentales qui ont été interprétées comme un reflet du rôle de contraintes phonotactiques ou syllabiques dans les processus de segmentation de la parole et qui pourraient différer de ce qu'on considère comme lié à l'utilisation d'indices probabilistes.

2.2.3.1. Les expériences de détection de syllabe

Mehler, Dommergues, Frauenfelder, & Segui (1981) ont mis en évidence un rôle de la syllabe dans les processus de traitement de la parole avec une tâche de détection de cibles. Celles-ci avaient soit une structure consonne-voyelle CV (par exemple /ba/) soit une structure

CVC (par exemple /bal/). Dans l'une des conditions expérimentales, la cible correspondait à la première syllabe du mot (/bal/ dans 'balcon', /ba/ dans 'balance'). Dans l'autre condition, il n'y avait correspondance qu'entre cette même cible et la séquence de phonèmes initiale (/bal/ dans 'balance', /ba/ dans 'balcon') mais aucune correspondance avec la structure syllabique du mot. Ainsi, le mot 'balcon' peut se découper en deux syllabes : /bal/ et /kɔ̃/. Si la cible indiquée à un participant est la séquence /bal/, il y a correspondance parfaite avec la syllabe initiale du mot. Par contre, si la cible est /ba/, ces deux séquences ne correspondent pas. Quoi qu'il en soit, dans tous les cas la chaîne de phonèmes initiale était appariée avec la cible. Qu'il y ait ou non correspondance syllabique, les participants pouvaient détecter la cible en position initiale du mot. Les auteurs observent des temps de détection plus courts lorsque cible et syllabe initiale correspondent que lorsque seule la séquence phonémique -indépendamment de sa structuration syllabique- est appariée avec la cible. Ce résultat est interprété par les auteurs comme un reflet du statut de la syllabe dans les processus d'accès au lexique : les représentations lexicales seraient stockées sous un format syllabique et le signal de parole devrait être analysé comme une suite de syllabes afin de l'apparier avec des représentations lexicales (Mehler et al., 1981; Segui et al., 1981). C'est également l'une des propositions avancées par Church (1987) qui, à partir d'une réflexion théorique sur le rôle des contraintes allophoniques et phonotactiques, estime que la syllabe serait une représentation intermédiaire adéquate pour faciliter la segmentation lexicale (*parsing*) et l'appariement (*matching*) avec les représentations lexicales.

Plus récemment, il a été montré (Cutler, Mehler, Norris, & Segui, 1986) que ces résultats ne se répliquent pas chez les locuteurs anglais. Dans une tâche identique, les locuteurs anglais présentent des temps de détection similaires qu'il y ait ou non correspondance entre la cible et la syllabe initiale du mot dans lequel elle apparaît. Ces données ont conduit les auteurs à modifier leur interprétation des résultats en insistant sur le rôle de la syllabe dans les processus de segmentation de la parole plutôt que dans les processus de classification du signal de parole préalables à un contact avec les représentations lexicales (Norris & Cutler, 1985; Segui, Dupoux, & Mehler, 1991). De même que les informations prosodiques pourraient être utilisées par le système de traitement de la parole en structurant les percepts linguistiques et / ou en guidant les processus d'activation lexicale, la structure syllabique des chaînes de phonèmes fournirait des indices pour la segmentation lexicale. Ces indices syllabiques seraient utilisés par les locuteurs français, langue dans laquelle la structure syllabique est relativement claire, alors que les locuteurs anglais utiliseraient plutôt des routines de segmentation fondées sur l'accent lexical.

Il a cependant été montré que les effets syllabiques obtenus avec cette tâche ne se répliquent pas lorsque les séquences porteuses sur lesquelles les participants doivent effectuer leurs jugements ne sont pas des mots (Content, Meunier, Frauenfelder, & Kearns, 1996) ; ce qui conduit les auteurs à envisager que la représentation syllabique impliquée dans l'émergence de ces effets ne peut être construite qu'à partir d'une représentation lexicale déjà constituée. On peut également objecter que les données obtenues en faveur d'un rôle de la syllabe dans les processus de segmentation de la parole l'ont été en demandant à des participants de manipuler explicitement des formes linguistiques syllabiques. En effet, la tâche des sujets consistait à donner une réponse concernant l'appariement éventuel entre la cible qui avait été déterminée au préalable et le stimulus qu'ils entendaient. Il est possible, du fait notamment de l'utilisation du concept de 'syllabe' dans l'apprentissage de la lecture mais aussi de la structure même de la langue française dans laquelle le phénomène d'ambisyllabicité est peu courant -alors qu'il est fréquent en anglais- que les locuteurs français éprouvent plus de facilités que les locuteurs anglais dans la manipulation de cibles syllabiques. Les effets observés pourraient alors s'expliquer par des phénomènes post-lexicaux impliquant l'émergence de stratégies de segmentation qui ne seraient pas réellement utilisées dans les processus de traitement de la parole mais se manifesteraient avec une tâche dans laquelle il apparaît que la syllabe est une unité adéquate et relativement maniable sur laquelle fonder les décisions nécessaires à l'accomplissement de l'expérience.

2.2.3.2. Syllabation et détection de phonèmes

Un rôle de la syllabe a cependant été mis en évidence avec un paradigme de détection de phonèmes, ce qui limite les possibilités d'intervention de stratégies de réponse reposant sur une segmentation non-écologique du signal perçu en syllabes. Vroomen & De Gelder (1999) ont présenté à des locuteurs néerlandais des phrases dans lesquelles les participants devaient détecter un phonème déterminé au préalable. Selon les conditions, le phonème-cible pouvait être prononcé en fin de syllabe comme dans la séquence :

'de.boo**t**.die.ge.zon.ke.nis'[16]

dans laquelle la cible est le phonème /t/. Dans cette situation, on observe qu'il y a correspondance entre la frontière syllabique qui sépare /but/ et /di/ et la frontière lexicale qui

[16] Les points introduits dans la représentation orthographique des phrases indiquent les frontières syllabiques.

sépare *boot* (‘bateau’) et *die* (‘qui’). Dans l’autre condition expérimentale, le phonème-cible apparaissait en position d’attaque syllabique comme dans l’énoncé :

‘de.boo.**t**is.ge.zon.ken’

On observe ici une discordance entre la frontière syllabique qui sépare /bu/ de /tiz/ et la frontière lexicale qui sépare *boot* (‘bateau’) et *is* (‘est’). Dans le premier cas, la consonne cible -qui est toujours une occlusive- est suivie d’une consonne ayant le même mode d’articulation (donc une occlusive). Lorsqu’il y a alignement des frontières, cette consonne est au contraire suivie d’une voyelle. Les participants devaient répondre le plus rapidement possible par l’appui sur un bouton réponse lorsque le phonème-cible était effectivement prononcé dans la phrase. Les auteurs observent des latences de détection plus courtes lorsque les frontières syllabique et lexicale sont alignées -c’est à dire lorsque le phonème-cible est en position de coda syllabique- que dans les séquences dans lesquelles il n’y a pas alignement entre ces frontières -quand le phonème à détecter est en attaque. Ils en concluent que c’est la non-correspondance entre découpage syllabique et frontière lexicale qui induit une difficulté à effectuer une segmentation lexicale correcte et nécessite des procédures de resyllabation afin d’être en mesure de localiser correctement cette frontière. Les locuteurs néerlandais auraient donc recours à une segmentation syllabique dans la tâche de détection de phonèmes.

Ce lien entre segmentation syllabique et détection phonémique s’expliquerait par un recours aux représentations lexicales pour extraire le phonème-cible. Il a en effet été mis en évidence que la tâche de détection phonémique peut être réalisée à partir de deux types de traitements. Il est possible de reconnaître un phonème par le biais d’une analyse purement prélexicale du signal : des processus d’appariement entre la représentation auditive du signal acoustique et les représentations phonémiques stockées en mémoire seraient alors mises en œuvre. Une autre procédure peut cependant être développée qui implique un recours aux représentations lexicales. Il est en effet possible de faire reposer la réponse de détection phonémique sur une préalable identification lexicale. L’existence de deux ‘voies’ possibles pour la réalisation de cette tâche a été mise en évidence dans des expériences dans lesquelles on fait varier la position du phonème à détecter et le statut lexical du stimulus qui le porte (Cutler, Butterfield & Williams, 1987 ; Frauenfelder, Segui & Dijkstra, 1990[17]). Dans ces travaux, on observe que les latences de détection de phonème varient à la fois en fonction du statut lexical mais aussi de la position dans le stimulus. Ces deux variables interagissent, l’effet du statut

[17] cf. Section 2.1 du **Erreur! Source du renvoi introuvable.**

lexical étant plus élevé lorsque le phonème-cible est prononcé en fin de stimulus qu'au début. Ainsi, si la tâche de détection de phonèmes utilisée par Vroomen & De Gelder (1999) peut reposer sur une identification préalable des mots présentés dans la phrase, toute difficulté à reconnaître les mots induira un délai dans les latences des réponses. C'est ce qui semble se passer lorsque les frontières syllabique et lexicale sont discordantes. La non-correspondance entre ces frontières induirait une difficulté transitoire à localiser la frontière lexicale et, par conséquent, à reconnaître le mot qui porte le phonème-cible. Ceci conduirait à accroître le temps nécessaire à la détection du phonème-cible par rapport à une situation dans laquelle les frontières sont alignées. Dans cette dernière condition, la frontière syllabique fournirait un indice de la présence d'une frontière lexicale. Ceci faciliterait la reconnaissance du mot et, de fait, la récupération de l'information phonémique nécessaire à la détection. On voit ici qu'une tâche n'impliquant pas la manipulation de cibles syllabiques permet de mettre en évidence des effets comportementaux interprétables en termes d'un recours à une segmentation syllabique de la chaîne parlée.

2.2.3.3. Régularités phonotactiques et *word-spotting*

McQueen (1998) a étudié le rôle des contraintes phonotactiques dans les processus de segmentation de la parole en avec des locuteurs dont la langue maternelle était également le néerlandais. Les contraintes phonotactiques régissent aussi bien l'agencement sériel des sons de parole que les sons qui sont prononçables par un locuteur natif. Celles que manipule McQueen (1998) portent sur les séquences de parole qui sont admises dans la langue en début de mot. Par exemple, en français, /tl/ et /dl/ sont des groupes de consonnes illégaux alors que /tr/ et /dr/ sont dits légaux. Les premiers n'apparaissent pas dans la langue en début de mot alors que les seconds sont attestés dans cette même position. McQueen (1998) distingue les phénomènes de syllabation de ceux liés aux régularités phonotactiques. En effet, les contraintes phonotactiques constituent seulement l'une des sources qui régissent la syllabation d'une séquence de phonèmes. Nous analyserons plus en détails dans le Chapitre 3 la distinction qui peut être faite entre régularités phonotactiques et syllabation. Nous nous contenterons de mentionner ici que les données présentées par Vroomen & De Gelder (1999) avec une tâche de détection de phonèmes et celles obtenues par McQueen (1998) en *word-spotting* peuvent tout à fait refléter l'utilisation d'informations similaires du fait du lien particulier qui existe entre les contraintes phonotactiques d'une langue et les procédures de syllabation des phonèmes dans la chaîne parlée. Par ailleurs, la dernière expérience présentée par McQueen (1998) afin de dissocier indices phonotactiques et syllabiques n'ayant pas donné de résultats concluants, il n'est pas possible de dire si les résultats

obtenus sont liés à des contraintes phonotactiques ou à des processus de syllabation de la chaîne parlée.

Afin de montrer que les contraintes phonotactiques constituent une classe d'informations utilisées dans le cadre des processus de segmentation de la parole en mots, McQueen (1998) reprend la tâche de *word-spotting* introduite par Cutler & Norris (1988) en manipulant le statut du groupe de consonnes médian. Les mots à détecter apparaissent soit en position initiale soit en position finale. Dans ces deux conditions, il fait varier la légalité phonotactique du groupe consonantique médian, celui-ci étant soit légal (/vr/, /dr/) soit illégal (*/mr/, */nr/). La tâche des sujets consiste à appuyer le plus rapidement possible sur un bouton réponse lorsqu'ils repèrent un monosyllabe de la langue en position initiale ou finale d'un non-mot de deux syllabes. La comparaison des données en position initiale et finale permet de contrebalancer la relation qui existe dans ces expériences entre légalité phonotactique du groupe de consonnes médian et alignement des frontières phonotactique et lexicale.

Tableau 2 : Statut du groupe de consonnes médian dans l'expérience de McQueen (1998) en fonction de l'alignement entre frontières phonotactique et lexicale.

	Non-alignement	**Alignement**
Position initiale	illégal /pɪlmrem/	légal /pɪlvrem/
Position finale	légal /fidrɒk/	illégal /fimrɒk/

Lorsque le mot à détecter est en position initiale, le groupe légal donne lieu à un non-alignement de la frontière phonotactique avec la frontière lexicale. Si le mot à détecter est *pil* ('pilule'), le stimulus /pɪlvrem/ -qui contient le groupe consonantique médian légal /vr/- donne lieu à une correspondance entre segmentation phonotactique /pil.vrem/ et segmentation lexicale /pil.vrem/. Lorsque le groupe médian est illégal (par exemple */mr/ dans /pɪlmrem/), il n'y a par contre plus correspondance entre segmentations phonotactique /pilm.rem/ et lexicale /pil.mrem/. Au contraire, si le mot à détecter est en position finale on observe une relation inverse entre légalité du groupe médian et alignement des frontières. Pour une détection du mot *rok* ('jupe'), la séquence /fimrɒk/ (dans laquelle */mr/ est illégal) donne lieu à un alignement de ces frontières -avec une segmentation phonotactique /fim.rɒk/- alors qu'un groupe de consonnes illégal (par exemple /dr/ dans /fidrɒk/) induit un non-alignement des découpages phonotactique (/fi.drɒk/) et lexical (/fid.rɒk/). Les informations présentées dans le Tableau 2 résument

l'agencement du lien entre légalité et alignement dans cette expérience. Les résultats obtenus lorsque les mots à détecter apparaissent en position initiale montrent que l'absence d'alignement entre les frontières phonotactique et lexicale -qui correspond à l'occurrence d'un groupe illégal en position médiane- donne lieu à des délais d'extraction du mot plus longs que la condition de non-alignement. Le mot *pil* est ainsi détecté plus lentement dans /pɪlmrem/ que dans /pilvrem/. On observe le même effet de l'alignement lorsque les mots à détecter sont en position finale ; la condition de non-alignement correspondant ici à un groupe consonantique médian légal. Dans cette condition, le mot *rok* est détecté plus lentement dans /fidrɒk/ que dans /fimrɒk/. McQueen (1998) en déduit un rôle des indices phonotactiques dans les processus de segmentation lexicale. Ces effets sont simulés (Norris, McQueen, Cutler, & Butterfield, 1997) dans une implémentation du modèle SHORTLIST (Norris, 1994) intégrant des processus de modification des niveaux d'activation en fonction des indices de segmentation phonotactiques.

2.2.3.4. Peut-on en déduire un rôle des régularités phonologiques dans les processus de segmentation lexicale ?

Les expériences réalisées par Vroomen & De Gelder (1999) et par McQueen (1998) aboutissent, avec des tâches qui n'induisent pas la manipulation directe de segments syllabiques de la part des auditeurs, à des résultats concordants en ce qui concerne un potentiel rôle des contraintes phonologiques (qu'elles soient d'origine phonotactique ou syllabique) dans les processus de segmentation de la parole en mots. Selon les auteurs, le système de reconnaissance de la parole utiliserait les connaissances intégrées au cours de l'apprentissage pour émettre des hypothèses sur de probables frontières lexicales ; ce qui faciliterait dans certains cas la reconnaissance des mots. On a vu cependant, dans le Chapitre 1, que des variables interprétées comme mettant en œuvre une catégorie particulière d'information (lexicale, phonologique ou fréquentielle) peuvent être facilement confondues en raison des particularités de la langue et des relations qui peuvent exister entre fréquence des mots, fréquence des diphones utilisés dans ces mots et régularités phonologiques. On peut par exemple envisager que, dans les expériences réalisées par Vroomen & De Gelder (1999) et McQueen (1998), il y ait une corrélation entre les modalités de la variable alignement (qui correspond à un contraste entre groupes de consonnes phonotactiquement légaux et illégaux ou entre séquences occlusive-voyelle et occlusive-occlusive) et la fréquence des séquences de phonèmes qui servent à la comparaison des modalités de cette variable.

Résumé

Nous avons vu, dans ce chapitre, deux approches complémentaires des processus de segmentation du signal de parole en mots. La première est fondée sur des processus d'accès au lexique aboutissant, par le biais de procédures de sélection ou de compétition lexicales, à la localisation de frontières entre les mots de la chaîne parlée. La seconde accentue le rôle d'informations prélexicales comme les régularités phonologiques (liées aussi bien à des indices segmentaux que suprasegmentaux tels que la prosodie et les contraintes phonotactiques) ou distributionnelles dans la mise en place d'hypothèses sur la localisation probable de frontières de mots. Les deux dernières études présentées ont conduit leurs auteurs respectifs à interpréter les résultats obtenus comme des preuves du recours à des connaissances sur les régularités phonotactiques ou syllabiques de la langue dans les processus de segmentation de la parole en mots. Il est cependant difficile de conclure de manière définitive à un rôle des informations phonologiques sans avoir une description adéquate des concepts de contraintes phonotactiques ou de syllabation et de leur expression dans la langue en termes de régularités probabilistes. L'objet du chapitre suivant est de présenter diverses approches des notions de légalité phonotactique et de syllabation tout en abordant la question du lien entre régularités phonologiques et fréquence.