

10.

LE RÔLE DE LA SYLLABE (*THE ROLL OF THE SILLY BALL*)

*Anne Cutler, James McQueen,
Dennis Norris et A. Somejuan*

In: *Les Langages du cerveau* (pp. 185-197). E. Dupoux (Ed.). Paris: Odile Jacob. (2002)

Quel est le rôle de la syllabe dans le traitement du langage parlé ? À notre avis, la réponse avancée par Jacques Mehler à cette question il y a à peu près vingt ans (voir Mehler, 1981) est incorrecte. Mais notre objectif n'est pas de nous acharner inutilement. Au contraire, nous voudrions apporter à Jacques la nouvelle réconfortante qu'il se trompait moins sur le rôle de la syllabe que nous ne le pensions. Selon des études récentes, il s'avère que la syllabe a bien un rôle universel à jouer dans le traitement de la parole, exactement comme Jacques Mehler l'a avancé il y a vingt ans. Mais ce n'est pas le rôle qu'il proposait alors.

L'étude expérimentale de la reconnaissance des mots parlés n'a qu'une trentaine d'années. Dès ses débuts, la syllabe a joué un rôle dans les débats théoriques (par exemple dans Savin et Bever, 1970) et Jacques Mehler était et reste encore l'un de ses plus solides défenseurs. La forme du débat a, bien entendu, quelque peu changé au cours de ces trente années, mais la syllabe, comme son fidèle champion, n'a pas baissé la garde.

POURQUOI LA SYLLABE N'EST PAS L'UNITÉ DE PERCEPTION

Nous essayons de démontrer ici que la syllabe a, en fait, deux rôles à jouer dans le traitement de la parole. Des études interlinguistiques suggèrent qu'un de ces rôles est spécifique à la langue alors que l'autre est universel. Mais aucun des deux ne correspond à la position adoptée par Mehler (1981), selon laquelle « la syllabe est

probablement le produit d'un mécanisme de segmentation opérant sur le signal acoustique. La syllabe sert ainsi à accéder au lexique » (p. 342). Selon cette position, les syllabes sont les « unités de perception » qui servent d'interface entre le signal de parole et le lexique mental. Cette théorie n'est plus tenable. Dans de nombreuses expériences effectuées dans les années 1970 et au début des années 1980 et qui comparaient les temps de détection de phonèmes et de syllabes, les syllabes étaient toujours détectées plus rapidement que les phonèmes (voir par exemple, Savin et Bever, 1970 ; Segui *et al.*, 1981 ; résumés dans Norris et Cutler, 1988). Cela pourrait laisser penser que, en accord avec la théorie selon laquelle les syllabes sont les représentations mentales utilisées lors de l'accès au lexique, la perception des syllabes se fait en premier et la détection des phonèmes dépend de la perception des syllabes. Cependant, comme l'ont fait observer Foss et Swinney (1973), il y a une faille dans cet argument : l'ordre d'identification des unités dans une tâche de laboratoire ne reflète pas nécessairement l'ordre dans lequel ces unités sont traitées dans la perception de la parole. Plus généralement, il n'est pas nécessaire de postuler que les unités que les sujets détectent dans les tâches de laboratoire correspondent directement aux représentations utilisées lors de l'accès au lexique (Cutler *et al.*, 1987). En outre, la supériorité de la syllabe observée dans les tâches d'identification s'est avérée être un artefact expérimental. Norris et Cutler (1988) ont demandé à des sujets de détecter soit des phonèmes cibles soit des syllabes cibles dans des listes de mots et de non-mots. Le matériel contenait des leurres : des phonèmes ou des syllabes qui étaient phonologiquement très similaires aux cibles. Contrairement à ce qui se passait dans les expériences antérieures, les sujets étaient donc obligés d'analyser entièrement chaque stimulus. Dans ces conditions, les phonèmes cibles étaient détectés plus rapidement que les syllabes cibles.

Des preuves plus convaincantes en faveur du rôle de la syllabe comme « unité de perception » ont été présentées par Mehler *et al.* (1981). Par exemple, les Français détectaient plus rapidement la séquence BA dans *balance* que dans *balcon* (où la cible correspond à la syllabification de *ba.lance*) et détectaient plus rapidement la séquence BAL dans *balcon* que dans *balance* (où la cible correspond à la syllabification de *balcon*). Bien que cette interaction croisée puisse servir à démontrer que le signal de parole est analysé en syllabes avant l'accès au lexique (comme l'affirmait Mehler, 1981, et Mehler *et al.*, 1981), il y a à cette observation une autre explication que nous lui préférons. Nous en discuterons plus loin. Pour le moment, notons seulement que, confrontés à exactement le même matériel français, ni les sujets anglais ni les Néerlandais n'ont montré le même effet syllabique (Cutler, 1997 ; Cutler *et al.*, 1983 ; 1986) et que cet effet n'a pu être reproduit quand des Français ont dû

détecter des cibles similaires dans des non-mots français (Frauenfelder et Content, 1999). Ces observations remettent en question l'hypothèse que la syllabe est l'unité prélexicale universelle.

Un autre problème majeur concernant cette hypothèse est que l'accès au lexique s'avère être un processus continu. La rapidité avec laquelle les mots parlés sont reconnus (Marslen-Wilson, 1973; Marslen-Wilson et Welsh, 1978) laisse à penser que l'accès au lexique commence dès les premiers sons d'un mot (Marslen-Wilson, 1987); l'initialisation de l'accès n'attend pas, par exemple, la fin de la première syllabe. Cette observation est incompatible avec tout modèle d'accès lexical comportant une stricte classification du signal sonore en syllabes qui intervient entre l'analyse acoustique et le lexique mental. Encore plus frappante est la démonstration que l'activation lexicale est modifiée par de subtils changements dans l'information acoustique-phonétique du signal de parole. Par exemple, les mots sont reconnus plus lentement quand ils contiennent une altération sub-phonémique que lorsqu'ils n'en contiennent pas (Marslen-Wilson et Warren, 1994; McQueen *et àL*, 1999; Streeter et Nigro, 1979; Whalen, 1984, 1991).

Ces études montrent que l'activation des représentations lexicales est modulée par des modifications du signal sonore qui sont non seulement subsyllabiques, mais aussi subphonémiques. Marslen-Wilson et Warren (1994) soutiennent donc que les représentations prélexicales consistent en traits phonologiques. Cependant, les données révélant une sensibilité aux modifications subphonémiques n'informent en rien de la nature des représentations prélexicales (McQueen *et àL*, 1999; Norris, McQueen et Cutler, 2000). Il est vrai que ces données excluent les modèles sériels pour lesquels l'information est transmise au niveau lexical en éléments discrets (que ces éléments soient phonémiques, syllabiques ou autres). Cependant, si l'information est transmise au lexique de manière continue, les modifications subphonémiques pourraient moduler l'activation lexicale quelle que soit la taille des unités qui codent cette information. Les données qui font état de la prise en compte continue de l'information au niveau lexical n'excluent donc que la forme la plus extrême du modèle syllabique, celle de la classification strictement sérielle en syllabes précédant l'accès au lexique.

Un autre argument contre la syllabe en tant qu'unité de perception provient de l'analyse de la structure phonologique des langues du monde. Dans certaines langues, la structure syllabique est très claire, dans d'autres, pas. Par exemple, il n'y a pas de consensus chez les anglophones (natifs) sur la syllabification de l'anglais (Treiman et Danis, 1988). Il n'y en a même pas sur la syllabification du français chez les francophones (Content *et àL*, 2000; Frauenfelder et Content, 1999). Les phonologues (comme Kahn, 1980) ont avancé que certaines consonnes en anglais (comme les consonnes intervocaliques

suivant une voyelle relâchée et précédant un schwa, comme le *IM* dans *balance*) sont ambisyllabiques, c'est-à-dire qu'elles appartiennent à la fois à la syllabe qui la précède et à celle qui la suit. Un modèle d'accès lexical fondé sur une classification en syllabes ne peut fonctionner de manière systématique dans les langues où l'ambisyllabité existe. À moins de postuler que la forme des représentations prélexicales varie d'une langue à l'autre, de sorte que seuls les locuteurs des langues possédant une structure syllabique totalement régulière accèdent au lexique par le biais d'unités syllabiques, l'ambisyllabité présente un élément de plus à l'encontre de l'hypothèse selon laquelle les syllabes sont les unités de perception. Si nous croyons qu'il existe d'importantes différences entre les langues dans le traitement de la parole en fonction de leurs structures phonologiques, nous pensons qu'il est peu vraisemblable que ces différences soient telles que l'accès des mots parlés repose sur des unités différentes dans des langues différentes.

La position extrême dont Mehler (1981) s'est fait l'avocat peut donc être rejetée. Remarquons cependant que nous sommes incapables d'exclure totalement la syllabe comme unité prélexicale. La nature des représentations mentales qui permettent d'accéder au lexique reste à déterminer : elles pourraient consister en gestes articulatoires (Liberman et Mattingly, 1985), en traits (Marslen-Wilson et Warren, 1994) en phonèmes (Fowler, 1984 ; Nearey, 1997) ou même en syllabes. Il est clair, cependant, que, si ces représentations s'avèrent être syllabiques, elles devront opérer de telle manière que l'information puisse être transmise de manière continue depuis elles jusqu'aux représentations lexicales, et d'une manière qui rende compte de l'ambisyllabité. Autrement dit, il ne peut y avoir de classification stricte du signal de parole en syllabes discrètes et indivisibles avant l'accès au lexique.

COMMENT LA SYLLABE EST UN MOT POSSIBLE

Comment alors peut-on expliquer les résultats de Mehler *et al.* (1981) ? Nous allons le faire dans un contexte qui n'était pas disponible au début des années 1980 : Shortlist, un modèle informatique de reconnaissance des mots parlés (Norris, 1994 ; Norris *et al.*, 1997). Shortlist ne répond pas à la question de l'unité de perception. L'input utilisé dans l'implémentation actuelle de Shortlist est une chaîne de phonèmes. Ce qui caractérise le modèle ne dépend cependant pas de la validité de cette présupposition. L'aspect que le modèle cherche à expliquer lui est postérieur dans la séquence de traitement : l'accès au lexique et la manière dont s'effectue la reconnaissance des mots parlés, en particulier la reconnaissance des mots dans le cas normal, celui de la parole continue. Cette reconnaissance est rendue difficile

pour deux raisons indépendantes. L'une est que la parole continue ne contient pas d'indices fiables indiquant les frontières de mots, ce qui oblige l'auditeur à segmenter le flux continu de sons produits par le locuteur afin d'identifier les mots présents dans le message du locuteur. La seconde raison est le caractère ambigu d'une grande partie du langage parlé : le début d'un mot donné est souvent consistant avec un grand nombre d'alternatives, et beaucoup de mots contiennent d'autres mots enchâssés. Cela complique encore le problème de la segmentation. En fait, il arrive souvent, pour un input donné, que de multiples analyses soient possibles, comme par exemple dans la séquence /ðasilbalzrol/.

Le modèle Shortlist a deux propriétés principales. La première est que seuls les mots les plus cohérents avec l'input sont activés (ils constituent la liste abrégée des meilleurs candidats, d'où le nom du modèle). La seconde propriété du modèle est que les mots candidats activés entrent en compétition (par l'intermédiaire de connexions inhibitrices). Cette compétition entre mots candidats apporte une solution au problème de la segmentation. Des mots comme *silly*, *balls* et *rott*, et d'autres comme *sill*, *syllabub* et *roe* sont activés en même temps que *syllable* quand le sujet entend /ðossilbolzroV/. Le signal sonore lui-même peut agir pour défavoriser certaines de ces alternatives (comme l'incompatibilité entre *syllabub* et la syllabe /bolz/; voir par exemple dans Connine *et al.*, 1997, et Marlsen-Wilson, Moss et van Halen, 1996, une discussion sur le rôle de ces incompatibilités lors de l'accès au lexique). La compétition lexicale fournit un mécanisme efficace par lequel les différents candidats peuvent être évalués les uns par rapport aux autres même en l'absence d'information claire dans le signal pour favoriser ou défavoriser des alternatives particulières (McQueen *et al.*, 1995). Le résultat de ce processus de compétition est une analyse lexicale correspondant à l'interprétation optimale de l'input. L'analyse *the silly balls roll* est meilleure que *the silly baUs roe* par exemple, puisque cette dernière ne prend pas en compte le /l/ final. Le candidat *roe* perdra dans sa compétition avec le candidat *roll* puisqu'il existe davantage d'éléments en faveur de *roll*.

Il y a cependant plus dans la segmentation de la parole que la compétition lexicale. S'il est vrai que la parole est continue et manque d'indices de frontière de mots fiables, il est incorrect de penser que le signal de parole ne contient aucun indice de frontière de mots. Un silence en est un clair indice : avant et après un énoncé, ou aux frontières des propositions à l'intérieur d'un énoncé. Il a été démontré que l'auditeur est sensible à ces indices et, en fait, à beaucoup d'autres que le signal de parole peut contenir (voir Norris *et al.*, 1997, pour une revue générale).

Dans Norris *et al.* (1997), nous avons cherché à unifier le fait que l'auditeur utilise ces différents indices de segmentation avec une

théorie de la reconnaissance des mots fondée sur la compétition. L'idée était que ces indices indiquent les sites probables de frontières de mots et que cette information pourrait servir à influencer le processus de compétition. Ce qu'il fallait donc, c'était un mécanisme par lequel chaque hypothèse lexicale activée puisse être évaluée à la lumière de l'information fournie par les indices de segmentation. Un candidat qui est parfaitement aligné avec une frontière de mot probable (c'est-à-dire un mot qui commence ou finit au point signalé comme une frontière probable) devrait avoir l'avantage sur celui qui n'est pas aligné avec une frontière. Mais qu'est-ce qui définit un mauvais alignement ? Le facteur décisif pour déterminer la valeur d'une segmentation lexicale correspondant à un input, c'est d'expliquer l'input en son entier de manière plausible. La contrainte d'alignement que nous avons donc proposée est la contrainte du mot possible (CMP). Si ce qui reste entre un mot candidat et une frontière de mot probable est un mot impossible, l'analyse qui comprend ce mot candidat est extrêmement peu plausible et le candidat en question est défavorisé (lors de l'implémentation de la CMP dans Shortlist, le niveau d'activation du mot est divisé par deux). Comme le silence à la fin du groupe /ôasilbalzr'ol/ indique une frontière de mot vraisemblable après le *IV* final, et que / tout seul n'est pas un mot possible en anglais, le candidat *roe* peut être pénalisé (l'énoncé *The silly batts roe l* est extrêmement peu plausible). Bien que, comme nous l'avons vu plus tôt, la compétition seule puisse favoriser *roU* aux dépens de *roe*, la CMP donne à *roe* un coup fatal (en le faisant descendre tout en bas de la liste abrégée des candidats).

Norris *et al.* (1997) ont présenté des données expérimentales qui jouent en faveur de la CMP. Des auditeurs devaient détecter des mots anglais enchâssés dans des séquences dénuées de sens. Ils ont trouvé beaucoup plus difficile de détecter un mot comme *apple* dans *fapple* que dans *vuffapple*. Selon la CMP, c'est parce que dans *fapple*, le mot *apple* n'est pas aligné avec la frontière indiquée par le silence au début de la séquence ; entre le mot et la frontière il y a une seule consonne *f* qui n'est pas un mot anglais possible. Dans *vuffapple*, en revanche, il y a une syllabe, *vuff*, entre *apple* et la frontière, et cette syllabe, sans être un vrai mot anglais, est un mot anglais possible. La pénalité de la CMP s'applique donc dans le premier cas mais pas dans le second. Les simulations avec Shortlist (également dans Norris *et al.*, 1997) ont montré non seulement que la CMP est nécessaire pour que le modèle puisse prédire ces données (et d'autres résultats sur la segmentation), mais aussi que les performances du modèle lors de la parole continue étaient meilleures avec la CMP que sans elle.

La CMP évalue les hypothèses lexicales activées en fonction des sites probables de frontières de mots quelle que soit l'origine des indices utilisés pour postuler ces sites. Les multiples indices de

segmentation présents dans le signal de parole peuvent donc tous influencer le processus de compétition par le même mécanisme. Il est évident que ces indices sont spécifiques à chaque langue. Ce qui correspond à une séquence de phonèmes phonotactiquement permise varie d'une langue à l'autre. L'harmonie des voyelles, par exemple, ne peut donner un indice de segmentation que dans une langue qui possède des contraintes d'harmonie vocalique (Suomi, McQueen et Cutler, 1997). D'autres indices qui sont nécessairement spécifiques à la langue sont ceux liés au rythme : différentes langues ont des propriétés rythmiques différentes. Un grand nombre de recherches comparatives suggèrent que l'auditeur utilise le rythme de sa langue native dans la segmentation de la parole (Cutler et Mehler, 1993 ; Cutler et Norris, 1988 ; Otake *et al.*, 1993).

OÙ LA SYLLABE JOUE UN RÔLE SPÉCIFIQUE À LA LANGUE

Avec cette perspective comparative en tête, nous pouvons maintenant considérer de nouveau le français et les résultats de détection de fragments de Mehler *et al.* (1981). Si les Français détectent plus rapidement BA dans ba.la.nce que dans *baLcon* et BAL dans *balcon* que dans *balance*, nous pensons que c'est parce qu'en français les syllabes peuvent servir à aider la segmentation de la parole lors de l'accès au lexique (voir la même idée dans Frauenfelder et Content, 1999, et Content *et al.*, 2000).

Le rôle de la syllabe dans la segmentation lexicale du français provient de son rôle dans la structure rythmique de cette langue. Si la syllabe ne semble jouer aucun rôle dans la segmentation lexicale qu'effectuent les locuteurs de certaines autres langues, c'est que, dans ces langues, le rythme n'est pas fondé sur la syllabe. En anglais et en néerlandais, par exemple, le rythme est fondé sur l'accent; en japonais, il est fondé sur une unité subsyllabique appelée la *more*. La structure rythmique, quelle que soit sa nature (spécifique au langage), donne des indices sur le site probable des frontières des mots, et ces indices, à leur tour, sont exploités par la CMP. Il est à noter que ces indices n'ont pas besoin d'être déterministes - peu importe si, par exemple, en français, la resyllabification fait que certaines limites de syllabes ne correspondent pas aux limites des mots, ou si, en anglais, certains mots commencent sur des syllabes non accentuées. L'accès au lexique est fondé sur l'arrivée continue d'informations sans être affecté par la structure rythmique. Cependant, les indices rythmiques, là où ils existent, se combinent avec les autres indices existants afin d'aider la CMP à éliminer certains des mots activés.

En d'autres termes, cette idée du traitement lexical est très différente de celle adoptée dans le débat sur les unités de perception

(Norris et Cutler, 1985). Aucune position n'est adoptée quant aux représentations de l'accès prélexical, ni n'a besoin de l'être, les informations données par le rythme de la langue (et par d'autres indices) indiquent simplement où, dans le signal, se trouvent vraisemblablement les frontières de mots. Ces informations sont ensuite utilisées par la CMP pour évaluer les hypothèses lexicales. L'affirmation qu'en français la structure syllabique marque les sites des frontières probables de mots ne dépend pas de présupposés particuliers sur les unités d'accès lexical. Si BA est plus facile à détecter dans *balance* que dans *balcon*, ce n'est donc pas parce que l'unité correspondant à /ha/ est extraite avant l'accès lexical qui reçoit *bculance* (mais pas *balcon*), mais parce que BA s'aligne parfaitement avec la frontière de mot probable indiquée par la structure de syllabe avant le /l/ dans *baJance*, alors qu'il ne s'aligne pas avec la frontière de mot probable indiquée par la structure de syllabe après le N dans *balcon*.

Nous préférons cette explication parce qu'elle a une validité à travers plusieurs langues. Elle explique pourquoi il apparaît que les locuteurs de langues différentes segmentent leur langue native selon des modes différents. Ce qui est commun à toutes les langues, c'est que le rythme de la langue d'origine donne des indices de segmentation. Cela explique aussi pourquoi les personnes qui parlent différentes langues segmentent aussi les langues non natives selon des modes différents. L'auditeur français segmente l'anglais syllabiquement alors que l'auditeur anglais ne segmente ni l'anglais ni le français syllabiquement (Cutler *et al.*, 1983, 1986). Comme l'auditeur anglais, l'auditeur néerlandais ne segmente pas le français syllabiquement (Cutler, 1997). L'auditeur français segmente aussi le japonais syllabiquement, mais pas l'auditeur japonais (Otake *et al.*, 1993). Enfin, même les bilingues français-anglais ne segmentent pas leurs deux langues selon le même mode. Cutler *et al.* (1989, 1992) ont trouvé que certains bilingues segmentaient le français syllabiquement, mais pas l'anglais ; chez les autres bilingues, nous n'avons pu mettre en évidence de stratégie syllabique dans aucune des deux langues, mais nous avons plutôt observé qu'ils segmentaient l'anglais (et pas le français) en utilisant une stratégie fondée sur l'accent, fondée sur le rythme de l'anglais. Une théorie dans laquelle la syllabe est l'« unité de perception » universelle ne peut expliquer cette variabilité interlinguistique.

COMMENT LA SYLLABE JOUE AUSSI UN RÔLE D'UNIVERSEL DU LANGAGE

Jacques Mehler serait peut-être déçu s'il devait se résoudre à ce que la syllabe ne joue qu'un rôle spécifique à la langue dans le traitement de la parole. Mais cette conclusion serait prématurée.

Nos récents travaux laissent à penser que la syllabe a aussi un rôle d'universel du langage.

La CMP dépend de l'activation simultanée de multiples mots candidats et d'indices sur les frontières probables de mots, mais elle dépend aussi d'un troisième facteur : une définition claire de ce qui compte comme un < mot possible », c'est-à-dire, une portion acceptable de discours entre un mot candidat et un site de frontière. Dans Norris *et al* (1997), nous avons comparé des portions consonantiques comme le *f* de *fapple* avec des portions syllabiques comme le *vuff* de *vuffapple*. Comme nous l'avons dit plus haut, la syllabe ici est clairement un mot anglais possible, mais pas la consonne. Dans l'application de la CMP, un mot est donc considéré comme un candidat acceptable si la portion du signal située entre la frontière de ce mot et une frontière de mot probable contient une voyelle (c'est-à-dire, si c'est une syllabe) et il est considéré comme un candidat non plausible si la portion du signal ne contient pas de voyelle.

La simple application de la CMP pourrait fonctionner dans toutes les langues. Mais est-ce le cas ? Les langues diffèrent dans ce qui constitue les mots possibles minimaux. Ainsi, en anglais, les syllabes consonne-voyelle (CV) dans lesquelles la voyelle est relâchée (par exemple, /vae/, *ItEf*) ne sont pas des mots bien formés, alors qu'en français, ce sont des mots bien formés (*va* et *thé* sont effectivement tous les deux des mots). Si l'implémentation actuelle de la CMP est correcte, les syllabes CV avec voyelles relâchées devraient être traitées comme des mots acceptables dans la segmentation de n'importe quelle langue. Dans une tâche de détection de mots, les mots devraient donc être plus faciles à repérer dans des contextes de voyelles ouvertes avec voyelles relâchées que dans des contextes avec consonnes uniques. Si cependant la CMP respecte le fait que les syllabes ouvertes avec voyelles relâchées violent les contraintes phonologiques de l'anglais, le repérage de mots devrait être aussi difficile en anglais dans les contextes avec voyelles relâchées que dans les contextes avec consonnes. Norris *et al.* (2000) ont testé cette prédiction. Les auditeurs anglais ont repéré *canal*, par exemple, plus rapidement et plus exactement dans un contexte CV avec voyelle relâchée (/zEcanal/) que dans un contexte avec consonne (/scanal/).

Une autre contrainte spécifique aux mots anglais est qu'une syllabe faible (contenant la voyelle réduite, le schwa) ne correspond pas à un mot de contenu bien formé. Les mots de la classe fermée des mots de fonction peuvent avoir un schwa comme voyelle unique (*the*, *a*, *of*, etc.) mais ceux de la classe ouverte des mots de contenu ont tous au moins une voyelle pleine. Norris *et al.* (2000) ont aussi testé si la CMP était sensible à cette contrainte sur le caractère bien formé des mots anglais. Les auditeurs détectaient des cibles comme *bell* plus vite dans /bElSaf/ que dans /bElS/. Cette différence était

équivalente à la différence démontrée par Noms *et al.* (1997), entre les contextes syllabique et consonantique pour le même ensemble de mots cibles où, par exemple, *bell* était plus facile à repérer dans des contextes avec des syllabes contenant des voyelles pleines (/bEISig/) que dans des contextes avec consonne (/bEIS/).

Les syllabes ouvertes contenant des voyelles relâchées ou un schwa ne semblent donc pas être traitées par la CMP comme des consonnes uniques. Cela laisse penser qu'il n'y a pas de correspondance entre ce qui constitue un mot possible dans une langue sur la base des contraintes phonologiques abstraites et ce qui constitue une portion de signal viable pour l'accès au lexique dans la reconnaissance de la parole continue. Cela a aussi été démontré dans une expérience de détection de mots en sesotho, une langue parlée dans le sud de l'Afrique. Les langues bantou comme le sesotho ont la contrainte phonologique que toute réalisation de surface d'un mot de contenu doit être au moins bisyllabique. Les monosyllabes ne sont donc pas des mots bien formés en sesotho. Cutler, Demuth et McQueen (en préparation) ont comparé l'aptitude des locuteurs de sesotho à détecter un mot comme *alafa* (prescrire) dans *halafa* (où le contexte d'une seule consonne *h* est un mot impossible) et dans *roalafa* (où le contexte monosyllabique *ro* n'est pas un mot, ni même un mot bien formé en sesotho). Les auditeurs étaient capables de repérer les mots dans les contextes monosyllabiques plus rapidement et avec moins d'erreurs que dans les contextes consonantiques. Il semblerait que le fait que *ro* n'est pas un mot possible dans le vocabulaire sesotho ne le rend pas inacceptable en tant que portion de l'analyse *ro + alafa*.

Tous ces résultats indiquent qu'à travers les langues du monde les segments de parole considérés comme découpables ne sont pas inférieurs à une syllabe. Autrement dit, la procédure de segmentation représentée par la CMP ne varie pas d'une langue à l'autre selon le caractère bien formé des mots dans chaque langue. Au contraire, le mécanisme de la CMP semble respecter une simple contrainte commune à toutes les langues : si la portion de signal entre une frontière de mot probable et la frontière d'un mot candidat est une syllabe, ce mot candidat est une partie viable de l'analyse ; si la portion de signal n'est pas syllabique (si c'est une seule consonne ou une suite de consonnes), le mot candidat ne fait pas partie d'une analyse plausible, et son activation est donc réduite. Ainsi, dans de nombreuses langues, y compris l'anglais et le sesotho, les auditeurs trouvent plus difficile de repérer des mots dans des contextes non syllabiques que dans des contextes syllabiques : en néerlandais, *lepel* (cuillère) est plus difficile à repérer dans /blepal/ que dans /salepal/ (McQueen et Cutler, 1998), il est à noter que les syllabes faibles ne sont pas non plus des mots de contenu possibles en néerlandais ; et en japonais, *ari* (fourmi) est plus difficile à repérer dans /rari/ que

dans /eari/ (McQueen, Otake et Cutler, sous presse). La taille ou la nature de la syllabe ne semblent pas importer : n'importe quelle syllabe fera l'affaire.

H semble donc que la syllabe a bien un rôle central à jouer dans le traitement de la parole. Les syllabes ne sont pas, comme Mehler (1981) le suggérait, des unités de perception que les auditeurs utiliseraient pour classer le signal sonore avant l'accès au lexique. Il apparaît plutôt que la syllabe est la mesure étalon grâce à laquelle les tentatives de segmentation du signal sonore sont jugées viables et non viables. Les syllabes forment des fragments acceptables lors de l'analyse lexicale du signal de parole, et non les séquences non syllabiques. *La syllabe est morte ; vive la syllabe* !*

REMERCIEMENTS

L'ordre de citation des auteurs (C, M, N, S) n'est pas arbitraire. Nous remercions Peter Gipson de nous avoir suggéré le titre, sous la forme d'une fort mémorable demande de tiré-à-part qui a longtemps orné le mur du bureau de Jacques.

RÉFÉRENCES

- CONNINE, C. M., TRONE, D., DEELMAN, T. & BLASKO, D. (1997), « Similarity mapping in spoken word recognition », *Journal of Memory and Language*, 37, 463-480.
- CONTENT, A., DUMAY, N. & FRAUENFELDER, U. H. (2000), « The rôle of syllable structure in lexical segmentation : Helping listeners avoid mondegreens », dans A. Cutler, J. M. McQueen & R. ZONDERVAN (Eds), *Proceedings of the SWAP Workshop*, Nimègue, Pays-Bas, Max Planck Institute for Psycholinguistics, p. 39-42.
- CUTLER, A. (1997), « The syllable's rôle in the segmentation of stress languages », *Language and Cognitive Processes*, 12, 839-845.
- CUTLER, A. & MEHLER, J. (1993), « The periodicity bias », *Journal of Phonetics*, 21, 103-108.
- CUTLER, A. & NORRIS, D. (1988), « The rôle of strong syllables in segmentation for lexical access », *Journal of Experimental Psychology : Human Perception and Performance*, 14, 113-121.
- CUTLER, A., DEMUTH, K. & MCQUEEN, J. M. (en préparation), Universality versus language-specificity in listening to speech.
- CUTLER, A., MEHLER, J., NORRIS, D. & SEGUI, J. (1983), « A language spécifique compréhension strategy », *Nature*, 304, 159-160.
- CUTLER, A., MEHLER, J., NORRIS, D. & SEGUI, J. (1986), « The syllable's differing rôle in the segmentation of French and English », *Journal of Memory and Language*, 25, 385-400.

* En français dans le texte (NiT).

- CUTLER, A., MEHLER, J., NORRIS, D. & SEGUI, J. (1987), «Phonème identification and the lexicon », *Cognitive Psychology*, 19, 141-177.
- CUTLER, A., MEHLER, J., NORRIS, D. & SEGUI, J. (1989), « Limits on bilingualism », *Nature*, 340, 229-230.
- CUTLER, A., MEHLER, J., NORRIS, D. & SEGUI, J. (1992), « The monolingual nature of speech segmentation by bilinguals », *Cognitive Psychology*, 24, 381-410.
- Foss, D. J. & SWBINEY, D. A. (1973), « On the psychological reality of the phonème : Perception, identification and consciousness », *Journal of Verbal Learning and Verbal Behavior*, 12, 246-257.
- FOWLER, C. A. (1984), « Segmentation of coarticulated speech in perception », *Perception & Psychophysics*, 36, 359-368.
- FRAUENFELDER, U. H. & CONTENT, A. (1999), « The rôle of the syllable in spoken word recognition : Access or segmentation? », dans *Actes des Journées d'Études linguistiques*, Nantes, Université de Nantes, p. 1-8.
- KAHN, D. (1980), « Syllable-structure specifications in phonological rules », dans M. Aronoff & M.-L. Kean (Eds), *Juncture*, Saratoga, CA, Anna Libri, p. 91-105.
- LIBERMAN, A. M. & MATTINGLY, I. G. (1985), « The motor theory of speech perception revised », *Cognition*, 21, 1-36.
- MARSLÉN-WILSON, W. D. (1973), « Linguistic structure and speech shadowing at very short latencies », *Nature*, 244, 522-523.
- MARSLÉN-WILSON, W. D. (1987), «Functional parallelism in spoken word-recognition », *Cognition*, 25, 71-102.
- MARSLÉN-WILSON, W., MOSS, H. E. & VAN HALEN, S. (1996), « Perceptual distance and competition in lexical access », *Journal of Experimental Psychology : Human Perception and Performance*, 22, 1376-1392.
- MARSLÉN-WILSON, W. D. & WELSH, A. (1978), « Processing interactions and lexical access during word recognition in continuous speech », *Cognitive Psychology*, 10, 29-63.
- MARSLÉN-WILSON, W. & WARREN, P. (1994), « Levels of perceptual representation and process in lexical access : Words, phonèmes, and features », *Psychological Review*, 101, 653-675.
- MCQUEEN, J. M. & CUTLER, A. (1998), « Spotting (différent types of) words in (différent types of) context », dans *Proceedings of the 5th International Conference on Spoken Language Processing*, vol. 6, Sydney, Australian Speech Science and Technology Association, p. 2791-2794.
- MCQUEEN, J. M., CUTLER, A., BRISCOE, T. & NORRIS, D. (1995), « Models of continuous speech recognition and the contents of the vocabulary », *Language and Cognitive Processes*, 10, 309-331.
- MCQUEEN, J. M., NORRIS, D. & CUTLER, A. (1999), « Lexical influence in phonetic decision making : Evidence from subcategorical mismatches », *Journal of Experimental Psychology : Human Perception and Performance*, 25, 1363-1389.
- MCQUEEN, J. M., OTAKE, T. & CUTLER, A. (sous presse), « Rhythmic cues and possible-word constraints in Japanese speech segmentation », *Journal of Memory and Language*.
- MEHLER, J. (1981), « The rôle of syllables in speech processing : Infant and adult data », *Philosophical Transactions of the Royal Society of London, Séries B : Biological Sciences*, 295, 333-352.

- MEHLER, J., DOMMERGUES, J.-Y., FRAUENFELDER, U. H. & SEGUI, J. (1981), « The syllable's rôle in speech segmentation », *Journal of Verbal Learning and Verbal Behavior*, 20, 298-305.
- NEAREY, T. (1997), « Speech perception as pattern recognition », *Journal of the Acoustical Society of America*, 101, 3241-3254.
- NORRIS, D. G. (1994), « Shortlist : A connectionist model of continuous speech recognition », *Cognition*, 52, 189-234.
- NORRIS, D. & CUTLER, A. (1985), «Juncture détection», *Linguistics*, 23, 689-705.
- NORRIS, D. & CUTLER, A. (1988), « The relative accessibility of phonèmes and syllables », *Perception and Psychophysics*, 43, 541-550.
- NORRIS, D., CUTLER, A., MCQUEEN, J. M., BUTTERFIELD, S. & KEARNS, R. (2000), « Language-universal constraints on the segmentation of English », dans A. Cutler, J. M. McQueen & R. Zondervan (Eds), *Proceedings of the SWAP Workshop*, Nimègue, Pays-Bas, Max Planck Institute for Psycholinguistics, p. 43-46.
- NORRIS, D., MCQUEEN, J. M., CUTLER, A. & BUTTERFIELD, S. (1997), « The possible-word constraint in the segmentation of continuous speech », *Cognitive Psychology*, 34, 191-243.
- NORRIS, D., MCQUEEN, J. M. & CUTLER, A. (2000), « Merging information in speech recognition : Feedback is never necessary », *Behavioral and Brain Sciences*, 23, 299-325.
- OTAKE, T., HATANO, G., CUTLER, A. & MEHLER, J. (1993), « Mora or syllable ? Speech segmentation in Japanese », *Journal of Memory and Language*, 32, 258-278'.
- SAVIN, H. B. & BEVER, T. G. (1970), « The non-perceptual reality of the phonème », *Journal of Verbal Learning and Verbal Behavior*, 9, 295-302.
- SEGUI, J., FRAUENFELDER, U. & MEHLER, J. (1981), « Phonème monitoring, syllable monitoring and lexical access », *British Journal of Psychology*, 72, 471-477.
- STREETER, L. A. & NIGRO, G. N. (1979), « The rôle of medial consonant transitions in word perception », *Journal of the Acoustical Society of America*, 65, 1533-1541.
- SUOMI, K., MCQUEEN, & J. M. CUTLER, A. (1997), «Vowel harmony and speech segmentation in Finnish », *Journal of Memory and Language*, 36, 422-444.
- TREIMAN, R. & DANIS, C (1988), «Syllabification of intervocalic consonants », *Journal of Memory and Language*, 27, 87-104.
- WHALEN, D. H. (1984), « Subcategorical phonetic mismatches slow phonetic judgments », *Perception & Psychophysics*, 35, 49-64.
- WHALEN, D. H. (1991), «Subcategorical phonetic mismatches and lexical access », *Perception & Psychophysics*, 50, 351-360.