

Intégration de sources multiples d'information dans l'identification lexicale et l'acquisition de nouvelles représentations lexicales

Odile Bagou*[°]; Ulrich H. Frauenfelder*

*Laboratoire de psycholinguistique expérimentale, Université de Genève, Suisse

[°] Institut d'orthophonie, Université de Neuchâtel, Suisse

Odile.Bagou@unine.ch

ABSTRACT

This study investigates how French listeners exploit phonological and phonetic cues in segmenting continuous speech into words. We examined how these listeners integrate multiple sources of information not only in lexical identification, using the word spotting task, but also in the storage of new lexical representations, using an artificial language learning task. Results showed that the specific segmentation cues examined had different weights in these two tasks. Syllable onsets, simultaneously cued by allophonic variations and phonotactics, played a predominant role in lexical identification while stress was a “last-resort” segmentation cue. In contrast, rhythmic information, particularly primary stress, played a greater role in lexical acquisition while syllable onsets were not used as segmentation cues. These results suggest that caution is required in relating results on lexical acquisition and lexical identification.

Keywords: Speech segmentation, lexical identification, lexical acquisition, rhythm, phonotactics, allophonic variations.

1. INTRODUCTION

Les processus d'identification lexicale d'une part, et d'acquisition lexicale d'autre part, résultent de la segmentation de la chaîne de parole. Lors de l'identification lexicale, la segmentation permet à l'auditeur de contacter le lexique mental dans lequel sont stockées les représentations lexicales qu'il doit identifier. Ainsi, les connaissances lexicales préalables des auditeurs sont, au moins partiellement, engagées dans la segmentation du flux de parole. De récents travaux ont même révélé sa supériorité par rapport aux indices sous-lexicaux [5]. Lors de l'acquisition de nouvelles représentations lexicales, en revanche, l'auditeur - qu'il s'agisse d'un adulte soumis à une langue artificielle [1] ou d'un nouveau-né [2]-, est confronté à un flux de parole continu dont il doit délimiter les unités linguistiques pertinentes, avec des connaissances lexicales qui, au mieux, sont en cours de construction voire absentes. Bien que la plupart des travaux envisagent tacitement que la contribution des indices sous-lexicaux est identique qu'il s'agisse de contacter un lexique mental préexistant ou de le construire, ce présupposé doit être confirmé. Ainsi, l'objectif de ce travail est de déterminer si la présence de connaissances lexicales modifie la contribution relative des informations fournies par la structure phonétique et phonologique de la parole.

Comme nous l'avons mentionné précédemment, les études abordant la segmentation sous l'angle de l'identification lexicale attestent de l'usage conjoint de connaissances lexicales - que nous n'aborderons pas dans ce travail - et d'indices sous-lexicaux fournis par la structure phonético-phonologique de la parole. En résumé, les variations allophoniques, les régularités phonotactiques, ou encore la structure métrique faciliteraient la localisation des frontières lexicales potentielles ([3], pour une revue). Lors de l'acquisition de la langue maternelle, les probabilités transitionnelles, les accents rythmiques et plus tardivement, les règles phonologiques, permettraient aux nouveau-nés de segmenter la chaîne de parole en unités lexicales [2]. De même, les quelques études abordant la segmentation grâce à un paradigme d'acquisition de langue artificielle [1][4], suggèrent que les probabilités transitionnelles et la structure métrique constituent des indices de segmentation efficaces chez l'adulte.

Dans ces études, chacune de ces informations sous-lexicales a été manipulée individuellement (voir cependant [5]) alors que, dans la parole naturelle, les frontières lexicales ne sont jamais marquées par une unique information. L'auditeur doit évaluer des indices complémentaires ou antagonistes afin de choisir la segmentation qui correspond le plus adéquatement à la structure du signal d'entrée. Dédire le fonctionnement du système de traitement en se basant sur des études explorant la contribution d'un unique indice à la fois peut donc amener à une vision sur-simplifiée, voire erronée, du phénomène. En effet, l'efficacité d'une information particulière peut être corroborée dans des situations expérimentales très contrôlées où les autres indices sont absents ou dégradés (e.g. parole bruitée [5]), mais nous ne pouvons néanmoins en déduire qu'elle serait assurée si des informations plus puissantes avaient été fournies simultanément par le signal de parole. Il est donc essentiel d'explorer comment le système de traitement intègre de multiples informations simultanément afin d'évaluer leur contribution relative dans le processus de segmentation lexicale.

Nos précédents travaux sur le français ont permis de valider la pertinence d'une heuristique de segmentation selon laquelle les débuts de syllabes, localisés grâce aux règles phonotactiques et aux variations allophoniques, sont utilisés pour accéder au lexique mental lors de l'identification lexicale (SOSH, [6]). Par ailleurs, nous avons montré que les prééminences rythmiques, et en particulier l'accent primaire final, constituaient des points d'ancrage privilégiés à partir desquels les unités lexicales

à acquérir peuvent se construire. Ainsi, la structure prosodique est efficacement utilisée par les auditeurs francophones pour acquérir de nouvelles représentations lexicales [7].

Le premier objectif de ce papier est d'explorer comment l'auditeur segmente lorsque le signal de parole fournit simultanément des indices d'alignements, syllabique et prosodique. Deux cas de figure peuvent être envisagés: soit les deux indices manipulés sont de puissance similaire, soit l'un des deux indices est plus efficace que l'autre, auquel cas une hiérarchie de puissance des indices de segmentation pourra être définie. Le second objectif est de savoir quelles relations peuvent être établies entre la segmentation accomplie lors de l'identification lexicale et la segmentation effectuée pour acquérir. En d'autres termes, l'efficacité des indices d'alignements syllabique et prosodique est-elle comparable dans les deux processus ou diffère-t-elle selon la tâche à accomplir?

Deux expériences ont donc été menées: la première sur l'identification lexicale (tâche de word spotting) et l'autre sur l'acquisition lexicale (apprentissage d'une langue artificielle). Par ailleurs, deux types d'indices de segmentation ont été manipulés: (1) les indices d'alignement syllabique, marqués simultanément par les règles phonotactiques et les variations allophoniques (A-Alignement syllabique, NA-Non-Alignement syllabique) et (2) les indices d'alignement prosodique, plus particulièrement l'accent primaire final (P-Présence ou NP-Absence d'un accent primaire sur la syllabe précédant le début de l'unité lexicale à identifier ou à acquérir).

2. EXPÉRIENCE 1 : IDENTIFICATION LEXICALE (WORD SPOTTING)

2.1. Méthode

Stimuli (Table 1)

Des mots-cibles, mono- ou bisyllabiques, ont été insérés en position médiane dans des séquences porteuses sans signification de structure de base [$\sigma 1$.mot. $\sigma 2$]. Pour tester l'effet de l'alignement syllabique initial sur le traitement lexical, les items-cibles étaient soit alignés (e.g. « lutte » dans /sin.lyt.bad/), soit non-alignés (e.g. « lutte » dans /si.plyt.bad/) avec un début de syllabe. Ainsi, les syllabes $\sigma 1$ et $\sigma 2$ de structure CVC ont été choisies pour induire ou non la formation d'un groupe consonantique avec le début de l'item enchâssé, lui-même débutant par une consonne liquide, /l/ ou /r/. Par ailleurs, pour tester l'effet de l'alignement prosodique, la syllabe précédant le mot à détecter portait (e.g. « lutte » dans /sin.#lyt.bad/) ou non (e.g. « lutte » dans /sin.lyt.bad/) une proéminence finale marquée par une augmentation de 30 % de la durée intrinsèque et par une f_0 accrue de 40Hz sur le noyau.

Table 1 : Manipulations des alignements initiaux, syllabique (.) et prosodique (#)

Condition d'alignement	Cible	Alignement Syllabique	Alignement Prosodique	Séquences porteuses
A-P	LUTTE	+	+	sin.# <u>lyt</u> .bad
A-NP		+	$\emptyset$	sin. <u>lyt</u> .bad
NA-NP		-	$\emptyset$	si.p <u>lyt</u> .bad
NA-P		-	-	si.#p <u>lyt</u> .bad

Participants

60 étudiants de l'université de Genève ont participé à l'expérience. Tous étaient suisses francophones et nul ne présentait de trouble auditif connu.

Procédure et tâche

Les participants étaient testés individuellement en chambre sourde. On leur expliquait qu'ils allaient être soumis à des séquences auditives sans signification contenant ou non un mot de la langue. Leur tâche consistait, le cas échéant, à détecter et répéter le mot-cible enchâssé le plus rapidement possible.

2.2. Résultats et discussion

Des Anovas à mesures répétées ont été menées avec les taux (figure 1) et les latences de détections correctes comme variables dépendantes. Elles révèlent que, lorsque la syllabation de la chaîne de parole donne lieu à un alignement des frontières lexicales et syllabiques initiales (A-NP), la reconnaissance lexicale est plus précise ($F_1(1,19)=11.7$; $p=.003$; $F_2(1,28)=2.4$; $p=.1$) et plus rapide ($F_1(1,19)=6.1$; $p=.02$; $F_2(1,28)=4.2$; $p=.05$) que lorsque la syllabation donne lieu à un non-alignement initial (NA-NP). Par ailleurs, les mots enchâssés ne sont ni plus précisément ($F_1(1,19)=2.8$; $p=.1$; $F_2 < 1$), ni plus rapidement ($F_s < 1$) détectés lorsque les frontières lexicales initiales sont indicées simultanément par des indices d'alignements syllabique et prosodique (A-NP vs. A-P). En revanche, lorsqu'un non-alignement syllabique est renforcé par des informations prosodiques de frontière (NA-P), une péjoration qualitative (%) des performances est observée ($F_1(1,19)=5$; $p=.04$; $F_2(1,28)=1.8$; $p=.2$).

En résumé, dans le processus d'identification lexicale, l'alignement syllabique indicé par les règles phonotactiques et les variations allophoniques de début de syllabe, constitue un indice de segmentation plus puissant que l'information rythmique, celle-ci étant néanmoins traitée.

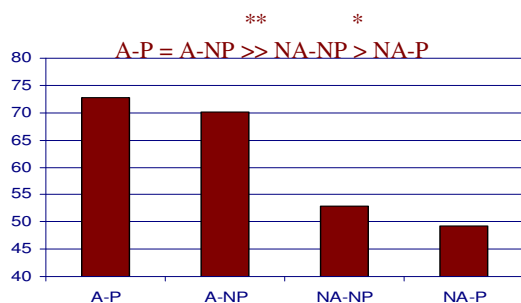


Figure 1: % d'identifications lexicales correctes

3. EXPERIENCE 2 : ACQUISITION LEXICALE (LANGUE ARTIFICIELLE)

3.1. Méthode

L'usage d'un paradigme d'acquisition de langue artificielle permet d'une part, de mettre des adultes dans une situation d'acquisition d'une langue dont ils n'ont aucune connaissance lexicale préalable et d'autre part, de manipuler les indices d'intérêt.

Mode de passation et tâche

La passation se déroule en deux temps: (1) une phase d'apprentissage et (2) une phase test, d'évaluation de la performance des participants. Lors de la première étape, les participants sont soumis auditivement à un extrait de 4 mn de la langue et doivent identifier et retenir les séquences de syllabes qui constituent, selon eux, un "mot" de la langue. Aucune information relative à la taille et à la structure des unités n'est donnée. Lors de la seconde étape, une tâche de préférence lexicale à choix forcé sans contrainte temporelle réelle, permet de vérifier l'intégration mnémorique des unités artificielles. Les participants sont soumis à des paires d'extraits de la langue correspondant soit à un "mot" et un "non-mot", soit à deux "non-mots". La tâche est de choisir celui des deux stimuli qui correspond ou ressemble à un "mot" de la langue artificielle préalablement apprise.

Construction de la langue de base

La langue artificielle de base résulte de la concaténation de 16 syllabes respectant les règles phonotactiques du français. Enregistrées par une locutrice francophone, les syllabes ont ensuite été re-synthétisées par la méthode PSOLA implémentée dans Praat à une fréquence de 210 Hz correspondant au fondamental moyen de la locutrice. Huit mots artificiels bisyllabiques débutant par une consonne liquide ont alors été créés par concaténation. Par ailleurs, 8 syllabes de type CVC dont les caractéristiques segmentales variaient selon la version de la langue, ont été adjointes à l'initiale des mots artificiels ainsi construits.

Conditions expérimentales

Quatre versions de la langue artificielle de structure [σ 1.mot1. σ 2.mot2. σ 3.mot3 ...] ont été construites. Les probabilités transitionnelles (PT) entre les mots artificiels et les syllabes adjointes σ variaient de .06 à .25 alors que les PT à l'intérieur des mots artificiels étaient égales à 1,

signalant ainsi que les syllabes apparaissent toujours de manière contiguë dans la langue. En outre, comme dans l'expérience précédente, les alignements initiaux, syllabique et prosodique, étaient manipulés (Table 2). Ainsi, les syllabes adjointes de structure CVC ont été choisies pour induire (versions NA) ou non (versions A) la formation d'un groupe consonantique avec la liquide initiale des mots artificiels. Par ailleurs, pour tester l'effet de l'alignement prosodique, les syllabes précédant le mot à acquérir et finales portaient (versions P) ou non (versions NP) une proéminence rythmique marquée par un allongement de 30% de leur durée intrinsèque et une élévation de f_0 de 20 Hz sur le noyau.

Table 2 : Manipulations des alignements initiaux, syllabique (.) et prosodique (en gras)

Versions LA	Alignement syllabique	Alignement prosodique	Extraits de la langue
A-P	+	+	<u>zes.luf.ged.kos.ryf.z</u> <u>aʃ.das.lët.mif</u>
A-NP	+	∅	<u>zes.luf.ged.kos.ryf.z</u> <u>aʃ.das.lët.mif</u>
NA-NP	-	∅	ti.glyb.daz.za.klët.mi f.ve.kluf.ged
NA-P	-	+	<u>ti.glyb.daz.za.klët.</u> <u>mif.ve.kluf.ged</u>

Construction des tests

Deux tests différents ont été construits selon que les mots artificiels étaient alignés ou non avec un début de syllabe dans la langue artificielle. Chaque test était constitué de 80 paires tests: 32 paires impliquant un mot (M) et un non-mot (NM) et 48 paires impliquant deux non-mots. Chacun des 8 mots artificiels était apparié avec les 32 non-mots partageant certaines syllabes avec les mots. Seules les analyses menées sur les paires M-NM sont présentées dans ce papier.

Participants

72 étudiants de l'université de Genève ont été répartis en 4 groupes expérimentaux de 18 participants chacun. Tous étaient suisses francophones et nul ne présentait de trouble auditif connu.

3.2. Résultats et discussion

Des ANOVAs à mesures répétées ont été réalisées - l'une considérant les participants (F1) et l'autre, les items (F2) comme facteur aléatoire - avec le taux de réponses correctes comme variable dépendante (Figure 2). Elles révèlent un effet principal significatif du facteur « version LA» ($F_1(3,68)=4.4$; $p=.007$; $F_2(3,21)=4.2$; $p=.02$), suggérant que la précision des réponses diffère selon la nature des indices dont disposent les participants pour extraire les mots artificiels. Plus précisément, les comparaisons multiples (tests post-hoc de Tukey) indiquent que l'information prosodique finale améliore considérablement les performances (NA-NP vs. NA-P :

$p=.04$) alors que l'information d'alignement syllabique ne constitue pas un indice efficace (A-NP vs. NA-NP : $p=1$).

En résumé, dans l'acquisition de nouvelles unités lexicales, la structure rythmique, et notamment l'accent primaire, constitue une source d'information plus puissante que l'alignement syllabique indicé par les règles phonotactiques et les variations allophoniques de début de syllabe.

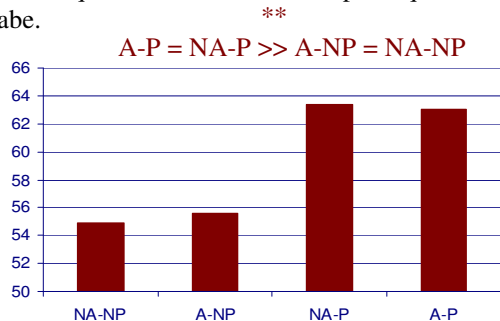


Figure 2: % d'identifications « lexicales » correctes après 4 mn d'apprentissage

4. DISCUSSION GENERALE

Dans ce travail, deux expériences ont été menées pour comparer la contribution relative de plusieurs indices de segmentation et établir des relations entre la segmentation accomplie lors de l'identification lexicale et lors de l'acquisition. Les résultats indiquent que, lors de l'identification lexicale, l'indice de segmentation le plus puissant est l'alignement syllabique alors que la structure rythmique n'accélère ni ne facilite l'accès au lexique mental. En revanche, lors de l'acquisition de nouvelles unités lexicales, la segmentation est principalement guidée par la structure rythmique et plus précisément par les accents primaires finaux, alors que l'information d'alignement syllabique ne facilite pas l'acquisition. Ainsi, les mécanismes sous-tendant les processus d'identification et d'acquisition ne sont pas strictement comparables. Nos résultats suggèrent que le poids des

sources d'information diffère, qu'il s'agisse de consulter un lexique mental préexistant ou de constituer de nouvelles représentations lexicales. En particulier, le rôle de l'information rythmique est considérablement accru lors de l'acquisition, ce qui pourrait être du à la saillance et à la résistance de ces caractéristiques par rapport aux autres informations linguistiques. En effet, la littérature rapporte que la segmentation d'une langue étrangère est influencée par les indices métriques pertinents dans la langue maternelle de l'auditeur et qu'aucune influence de la structure métrique de la langue cible n'est observée, même lorsque les participants sont d'excellents bilingues [8]. En revanche, des auditeurs soumis à une langue étrangère pourraient profiter de leur connaissance des règles phonologiques de la langue cible pour segmenter le flux de parole en mots [9]. Ainsi, les caractéristiques prosodiques seraient plus résistantes au changement que les connaissances phonologiques, ce qui pourrait expliquer pourquoi leur rôle est renforcé lors de l'acquisition.

En conclusion, ce travail suggère que les inférences formulées quant au fonctionnement du système d'acquisition à partir de données obtenues dans des tâches d'identification lexicale et inversement, requièrent la plus grande prudence. En effet, la contribution relative des indices sous-lexicaux varie en fonction de la tâche. Une explication possible pourrait être que la présence de connaissances lexicales modifie la hiérarchie des indices de segmentation sous-lexicaux. Ainsi, il serait intéressant d'explorer si et comment les informations lexicales, construites sur la base d'informations phonétiques et prosodiques, supplantent progressivement ces indices sous-lexicaux jusqu'à les rendre optionnels dans le processus de segmentation. En effet, au cours de l'acquisition, la hiérarchie des indices de segmentation se modifie probablement progressivement au fur et à mesure que la connaissance lexicale s'accroît.

framework. *Journal of Experimental Psychology: General*, 134: 477-500, 2005.

BIBLIOGRAPHIE

- [1] Saffran, J. R., Newport, E. L., & Aslin, R. N. Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35: 606-621, 1996.
- [2] Saffran, J. R., Aslin, R. N., & Newport, E. L. Statistical learning by 8-month-old infants. *Science*, 274: 1926-1928, 1996.
- [3] Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. Leading up the lexical garden-path: Segmentation and ambiguity in spoken-word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28: 218-244, 2002.
- [4] Banel, M. H., Frauenfelder, U. H. & Perruchet, P. Contribution des indices métriques à l'apprentissage d'un langage artificiel. In *JEP'98*, 29-32, 1998.
- [5] Mattys, S. L., White, L., & Melhorn, J. F. Integration of multiple speech segmentation cues: A hierarchical
- [6] Content, A., Kearns, R. K. & Frauenfelder, U.H. "Boundaries versus onsets in syllabic segmentation", *Journal of Memory and Language*, 45: 177-199, 2001.
- [7] Bagou, O., Frauenfelder, U.H. Stratégie de segmentation prosodique: rôle des proéminences initiales et finales dans l'acquisition d'une langue artificielle, In *JEP'06*, 2006.
- [8] Cutler, A., Mehler, J., Norris, D., & Segui, J. The monolingual nature of speech segmentation by bilinguals, *Cognitive Psychology*, 24(3): 381-410, 1992.
- [9] Weber, A. The role of phonotactics in the segmentation of native and non-native continuous speech. In *Proceedings of the Workshop on Spoken Access Processes*, 143-146, 2000.