Stratégie de segmentation prosodique : rôle des proéminences initiales et finales dans l'acquisition d'une langue artificielle

Odile Bagou et Ulrich H. Frauenfelder

Laboratoire de Psycholinguistique Expérimentale, FPSE, 40 bd du Pont D'Arve, 1211 Genève, Suisse
Odile.Bagou@pse.unige.ch

ABSTRACT

Language acquisition in infants and adults depends upon both the segmentation of the words in the speech chain and the extraction of language-specific regularities, particularly prosodic regularities. Two experiments investigate whether prosodic information provided by accented syllables located at the beginning and at the end of prosodic words is used by adult French learners to segment an artificial language. The results allow us to define a prosodic strategy of segmentation: (1) prominence in word final position are used to hypothesize final boundaries; and (2) cues in word initial position can be used if and only if a primary final prominence is present.

1. INTRODUCTION

Aucun indice acoustique ne marquant systématiquement les frontières lexicales dans la parole continue, la question de savoir quelles informations sont exploitées pour accomplir le processus de segmentation reste entière. Si l'on adopte une approche sous-lexicale de la segmentation, deux principales pistes de recherche ont été explorées dans la littérature. Alors que certains ont mis en évidence le rôle prépondérant des régularités phonotactiques et probabilités transitionnelles dans la segmentation [8][10], d'autres ont suggéré que la segmentation dépendait essentiellement de l'information prosodique [7]. La stratégie de segmentation métrique (MSS) [4] suggère que, dans les langues accentuelles telles que l'anglais, les syllabes fortes indiquent les frontières lexicales initiales potentielles. 87% des mots de contenu étant accentués sur la syllabe initiale en anglais, l'efficacité de cette stratégie est assurée [3]. Cependant, la MSS n'est pas directement applicable à une autre langue. Elle doit s'adapter à la structure métrique de la langue considérée. Le but de cet article est d'expliquer comment l'information prosodique guide la segmentation en français, une langue dans laquelle le domaine de la proéminence est supérieur au mot [5].

Les études récentes sur la parole spontanée admettent que le français possède un système accentuel dual. Ainsi, les accents finaux et initiaux coexisteraient [5]. De par sa nature obligatoire et sa place systématique sur la dernière syllabe pleine des mots de contenu, l'accent final remplirait principalement une fonction de démarcation à droite de l'unité [1]. Frappant les syllabes initiales des mots [5], l'accent secondaire quant à lui, remplirait une fonction démarcative de marquage gauche de l'unité. Ainsi, la coexistence des proéminences initiale et finale servirait à renforcer la cohésion d'une unité et pourrait être utilisée dans la segmentation [6].

Le principal objectif de ce travail est de tester la réalité cognitive de cette fonction démarcative supposée des proéminences initiales et finales dans la segmentation de la parole continue. Plus spécifiquement, puisque l'acquisition requiert une segmentation préalable du signal de parole, nous proposons de préciser la contribution des différentes proéminences, initiale et finale, dans l'acquisition d'une langue artificielle (LA). Enfin, puisque le marquage acoustique des proéminences finales est pluri paramétrique, plusieurs indices pourraient contribuer à la segmentation. L'accent final étant marqué, entre autres, par un allongement final et des variations mélodiques, nous proposons d'explorer la contribution relative de ces indices mélodiques et/ou temporels dans la segmentation de la parole continue.

2. EXPERIENCE 1 : PROEMINENCES FINALES

2.1. Méthode expérimentale

L'usage d'un paradigme d'acquisition de langue artificielle, peu usité en psycholinguistique [1,2,10], nous semblait parfaitement répondre à nos attentes: d'une part, parce qu'il permet de mettre des adultes disposant de connaissances linguistiques préalables dans une situation d'acquisition et d'autre part, parce qu'il permet la manipulation des facteurs 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, la langue est soumise auditivement aux participants. Pendant ce bain de langage, la tâche est de retenir les séquences de syllabes qui constituent 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 mnémonique des unités artificielles. Les participants sont soumis à des paires présentées auditivement, constituées de deux 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 plus à une "unité lexicale" de la langue artificielle apprise préalablement.

Construction de la langue de base

La langue artificielle de base est le résultat de la concaténation de 18 syllabes respectant les règles

phonotactiques du français. Enregistrées par un locuteur francophone suisse romand, les syllabes sont ensuite re-synthétisées selon la méthode PSOLA. Ainsi, les variations de fréquence fondamentale ont été harmonisées sur chacune des syllabes pour obtenir une fréquence uniforme de 110 Hz, correspondant au fondamental moyen du locuteur. Huit mots artificiels, 4 bi- et 4 trisyllabiques, ont alors été créés par concaténation. Enfin, cent occurrences de chacun des huit mots artificiels ont été concaténées dans un ordre semi-aléatoire en une séquence de parole continue, sans pause ni indice signalant la présence d'une frontière. Le même ordre de présentation était utilisé dans les différentes versions de la LA.

Conditions expérimentales

Quatre versions de la LA ont été construites : une LA de base (A) sans indices prosodiques de frontière : les auditeurs pouvaient inférer les unités en calculant des statistiques de co-occurrence des syllabes adjacentes (Probabilités transitionnelles PT) ; et 3 LA dont les unités portaient une proéminence finale (B,C,D). Dans ces versions prosodiquement marquées, au moins un indice acoustique caractéristique de la proéminence finale était fourni en plus de l'information statistique. Les indices mélodiques et temporels étaient tous deux disponibles dans la version D, alors que dans les versions B et C, seul l'un des deux indices était fourni : l'indice d'allongement (B) ou l'indice mélodique (C).
Les manipulations acoustiques ont été réalisées par le logiciel Praat, lequel nous a permis d'allonger la dernière syllabe de chaque unité de 30% de sa durée intrinsèque (B et D) et/ou d'augmenter la hauteur mélodique de la dernière syllabe de 110 à 130 Hz (C et D).

Construction du test

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. Les paires tests étaient construites en appariant chacun des 8 mots artificiels avec les 32 non mots partageant certaines caractéristiques des mots. Seules les analyses menées sur les paires M-NM sont présentées dans cet article.

Participants

96 suisses romands ont été répartis en 8 groupes expérimentaux de 12 participants chacun : 4 groupes tests appariés à 4 groupes contrôles, respectivement exposés ou non à la langue artificielle durant la phase d'apprentissage. De plus, l'ordre de présentation des paires testées était contrebalancé pour annuler des artefacts potentiels d'apprentissage par la tâche. Ainsi, la moitié des participants d'un groupe répondait à la première partie du test avant la deuxième, et inversement pour l'autre moitié de participants.

2.2. Résultats

Des t-tests ont été calculés afin de savoir si les participants obtenaient une performance supérieure au seuil de hasard (50%) dans chaque condition. Tous sont significatifs sauf pour le groupe contrôle *(t(47)=-.25, p=.8)*, ce que nous attendions puisque ces participants n'avaient pas subi la phase d'apprentissage préalable.
Des ANOVAs à mesures répétées ont été réalisées - l'une considérant les participants (F_1) et l'autre, les items (F_2) comme variable aléatoire - avec le taux de réponses correctes comme variable dépendante. Les analyses contenaient un facteur intra-sujets, le type de paire, et trois facteurs inter-sujets à savoir, la condition expérimentale, l'ordre de présentation des blocs dans le test et le groupe expérimental. Les résultats sont présentés dans la Table 1.

Table 1 : Taux de réponses correctes moyen (%) selon de groupe et les indices disponibles lors de l'acquisition

Versions LA	**Indices disponibles lors de l'acquisition**			
	A (PT)	B (PT + All)	C (PT + f0)	D (PT+ All. +f0)
Test	64,6	82,0	89,8	85,4
Contrôle	51,8	47,9	51,3	47,7
D	**12.8***	**34.1****	**38.5****	**37.7****

La significativité de l'effet du groupe (contrôle vs. test) suggère que la phase d'apprentissage est nécessaire à l'acquisition et à la rétention des mots artificiels ($F_1(1,80)= 274.2$; $p<.0001$; $F_2(1,6)=112.2$; $p<.0001$). Une amélioration de la performance (de 30% en moyenne) due à l'apprentissage préalable apparaît et ce, quelle que soit la condition expérimentale, i.e. quels que soient les indices disponibles dans la langue apprise.
De plus, l'effet significatif de la condition expérimentale suggère que les participants soumis à une langue dont les frontières finales d'unité sont indicées par des informations prosodiques obtiennent de meilleures performances que ceux qui ne disposent que d'indices statistiques de co-occurrence des syllabes pour intégrer les mots ($F_1(3,80)=7.7$; $p=.0001$; $F_2(3,18)=5$; $p=.01$). Ainsi, ces résultats indiquent que l'allongement final et/ou les variations mélodiques permettent une segmentation plus précise de la chaîne de parole que le calcul statistique seul.
Les comparaisons multiples (test de Tukey) indiquent également que la différence d'environ 6% observée entre les performances des participants soumis à la condition C (f0) et celles des participants soumis à la condition B (all.) tend à la significativité ($p=.1$). Ainsi, les variations mélodiques pourraient constituer un indice de frontière plus puissant que l'allongement final.
Par ailleurs, le cumul d'informations acoustiques (D : f0+allongement final) n'améliore pas significativement la performance. En effet, aucune différence n'est avérée entre les conditions dont les mots ne portent qu'un seul indice acoustique d'allongement (B) ou de f0 (C) et la condition D (respectivement, $p = .9$; $p = .4$).
Enfin, l'effet de la condition expérimentale interagit significativement avec le groupe expérimental ($F_1(3,80)=10.7$; $p<.0001$; $F_2(3,18)=7.8$; $p=.002$). Ainsi, la différence de performance entre les groupes contrôle et test est plus conséquente lorsque les indices prosodiques de frontière sont présents que lorsque seules les informations

statistiques sont disponibles (37 % en moyenne contre 12.8% dans la condition sans indice prosodique).
En résumé, qu'elle soit marquée par un allongement ou des variations mélodiques, la présence d'une proéminence finale permet un apprentissage plus efficace. Cependant, la redondance du marquage n'apporte aucune information supplémentaire.

3. EXPERIENCE 2 : PROEMINENCES INITIALES

3.1. Méthode expérimentale

La méthode générale de passation est identique à celle de l'expérience précédente (voir chap.2.1.). Seules les caractéristiques de la langue diffèrent.

Construction de la langue de base

La langue artificielle de base est constituée de 24 syllabes respectant les règles phonotactiques du français. Enregistrées par une locutrice francophone suisse romande, les syllabes sont ensuite re-synthétisées grâce au logiciel Praat selon la méthode PSOLA. Ainsi, les variations de fréquence fondamentale sont harmonisées pour obtenir une fréquence uniforme de 210 Hz, correspondant au fondamental moyen de la locutrice. Huit mots artificiels trisyllabiques, ont été créés par concaténation. Enfin, 96 occurrences de chacun des huit mots ont été concaténées dans un ordre semi-aléatoire en une séquence de parole continue, sans pause ni indice signalant la présence d'une frontière.

Conditions expérimentales

Trois versions de la LA ont été générées : une LA de base sans indices prosodiques de frontière (« plate »): les frontières d'unités pouvaient être inférées en considérant que des probabilités transitionnelles faibles entre les syllabes signalent la présence d'une frontière. Par ailleurs, 2 LA dont les unités portaient une proéminence initiale accompagnée ou non d'une proéminence finale ont été construites. Dans la version « Ai », seule la proéminence initiale était présente et marquée par une fréquence fondamentale variant de 230Hz au début de syllabe, à 250Hz au centre du noyau vocalique [11]. Dans la version « Arc », les unités portaient des proéminences initiales et finales. Pour le marquage de la proéminence finale, la fréquence fondamentale moyenne de 210Hz était élevée à 250Hz au début de la syllabe pour atteindre 270 Hz, au centre du noyau vocalique. De plus, la dernière syllabe de chaque unité était allongée de 30% de sa durée intrinsèque. Pour le marquage de la proéminence initiale, les caractéristiques étaient similaires à celles de la langue « Ai ». Les manipulations acoustiques ont été réalisées grâce au logiciel Praat.

Participants

51 francophones monolingues, étudiants en psychologie à l'université de Genève, ont été répartis en 3 groupes expérimentaux de 17 participants chacun.

3.2. Résultats

Des t-tests ont été calculés afin de savoir si les participants obtenaient une performance supérieure au seuil de hasard dans chaque condition. Tous sont significatifs, ce qui suggère que les participants n'ont pas répondu au hasard (50%).
Des ANOVAs à mesures répétées ont été réalisées - l'une considérant les participants (F_1) et l'autre, les items (F_2) comme variable aléatoire - avec le taux de réponses correctes comme variable dépendante. Les analyses contenaient un facteur intra-sujets, le type de paire, et un facteur inter-sujets, la condition expérimentale. Un effet principal de la condition expérimentale significatif apparait ($F_1(2,48)=6.2$; $p=.004$; $F_2(2,14)=9.5$; $p=.002$) (Figure 1).

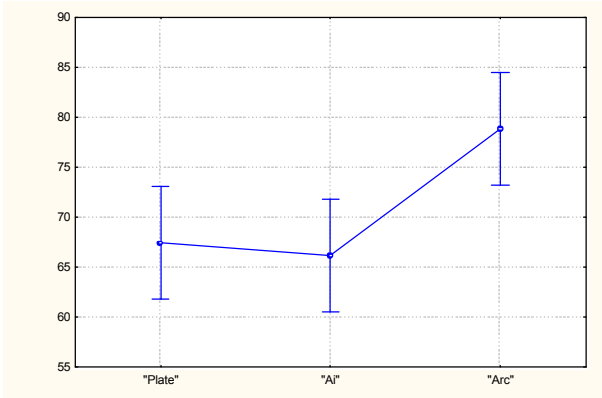

Figure 1 : Taux de réponses correctes moyens (%) selon le type de proéminence disponible durant l'apprentissage

Les comparaisons multiples (test de Tukey) confirment que la différence de performance entre les versions « plate » et « Arc » de la langue est significative ($p = .02$). Ainsi, le bornage bipolaire des unités artificielles fournit un indice pertinent de segmentation et favorise l'acquisition des unités. De même, la différence de performance entre les versions « Ai » et « Arc » est significative (p=.007), ce qui suggère que la proéminence initiale seule n'est pas aussi efficace que l'arc accentuel.
Par ailleurs, la performance n'est pas améliorée par la présence unique d'une proéminence initiale (« plate » vs. « Ai »: $p=.9$). Ainsi, l'accent initial ne permettrait pas de segmenter plus efficacement le signal de parole.

4. DISCUSSION GENERALE

Les résultats des deux expériences indiquent que l'information portée par les syllabes accentuées facilite l'acquisition d'une nouvelle langue. Dans les versions A (exp. 1) et « plate » (exp. 2), les participants ne disposaient que de l'information phonotactique et distributionnelle (PT). Leur faible performance corrobore le fait que ce type d'indice n'est pas suffisant pour segmenter idéalement le signal de parole. En revanche, la présence d'informations prosodiques sur les syllabes finales facilite l'acquisition de la langue : l'expérience 1 atteste de la pertinence des proéminences finales dans l'acquisition et l'expérience 2 indique que le bornage bipolaire des unités facilite leur intégration mnémonique.

L'expérience 1 a été également menée pour explorer la contribution relative de différents indices typiques des proéminences finales dans la segmentation. Puisque les participants de la condition B obtiennent des performances comparables à ceux de Banel et al. [2], l'effet de l'allongement final a été répliqué. Etonnement, les variations mélodiques finales seules induisent des réponses plus précises que l'allongement final seul. Ce résultat infirme les conclusions de Rietveld [9], lequel suggère que l'allongement final est l'indice le plus efficace pour segmenter des phrases dont la frontière est ambiguë. Cependant, dans notre LA, l'occurrence régulière d'un contour continuatif toutes les 2 ou 3 syllabes a pu faciliter la segmentation puisque les auditeurs pouvaient supposer qu'ils écoutaient une liste d'items plutôt que des phrases. Ainsi, cette prééminence des variations mélodiques dans la segmentation doit être répliquée avant d'être interprétée. Par ailleurs, on pourrait supposer que cette amélioration de la performance est le fait d'une impression subjective d'allongement due à la présence d'une montée mélodique. Cependant, si la meilleure performance était simplement due à un cumul « virtuel » d'indices, alors la performance aurait du être améliorée lorsque les indices étaient effectivement cumulés (version D), ce qui n'est pas le cas.

Alors que le bornage bipolaire (« arc ») permet une meilleure rétention des unités artificielles, la présence unique de la proéminence initiale (« Ai ») n'améliore pas l'acquisition dans l'expérience 2. Puisqu'en français, les constituants prosodiques ont la « tête » à droite, il était raisonnable de penser que les participants auraient attribué un statut de proéminence primaire finale à cette syllabe saillante. Ainsi, les frontières signalées par l'information prosodique auraient été différentes de celles suggérées par le calcul statistique (PT). Le fait que la performance dans la version « Ai » soit comparable à celle de la version « plate » aurait donc suggéré que les participants avaient privilégié les frontières indicées par les PT. Toutefois, nous avons montré que l'acquisition était facilitée par la présence d'une proéminence finale (exp.1), ce qui nous invite à rejeter l'idée que cette syllabe saillante ait été interprétée comme une proéminence finale.

De plus, puisque les patrons appliqués à nos unités artificielles ont été observés en début de mot de contenu dans des productions naturelles [11], il est probable qu'ils aient été interprétés comme des contours typiques de la proéminence initiale. En effet, le système pourrait être capable de détecter des motifs prosodiques fréquemment rencontrés aux frontières d'unités, ce qui lui éviterait de confondre des proéminences initiales et finales. Dans ce cas, nos résultats indiquent que la proéminence initiale n'a pas été utilisée pour segmenter les unités artificielles, ce qui suggère que les proéminences initiales ne rempliraient pas le rôle démarcatif qu'on leur attribue.

Cependant, cette absence d'effet pourrait être due à la spécificité du matériel. Puisque la longueur syllabique de toutes les unités artificielles était identique, les participants ont pu appliquer une stratégie triviale de dénombrement de syllabes, plutôt que d'utiliser l'information prosodique pour découvrir les mots de la langue. Cette explication est toutefois peu probable puisque nous observons, avec un matériel comparable, une amélioration nette de la performance due à la présence conjointe de proéminences initiales et finales (« Arc »).

Par ailleurs, les unités de notre LA sont des pseudo-mots de contenu qu'aucun pseudo-mot grammatical ne sépare. Or Welby [11] montre que la montée mélodique initiale caractérise une frontière entre mot grammatica1 et mot de contenu. Ainsi, la montée mélodique initiale n'aurait pu remplir sa fonction, ce qui explique qu'elle n'ait pu intervenir dans l'acquisition.

Pour conclure, la stratégie de segmentation prosodique serait principalement basée sur les proéminences finales dont la présence constituerait une condition nécessaire à l'usage éventuel des proéminences initiales. Reste à préciser si leur fonction est de signaler la fin de l'unité en cours de traitement ou, plus plausiblement, d'attirer l'attention vers le début de l'unité suivante.

5- BIBLIOGRAPHIE

[1] Bagou, O., Fougeron, C., Frauenfelder, U.H. 2002. Contribution of Prosody to the Segmentation and Storage of "Words" in the Acquisition of a New Mini-Language, *SP2002*, Aix, France.

[2] Banel, M. H., Frauenfelder, U. H. and Perruchet, P. 1998. Contribution des indices métriques à l'apprentissage d'un langage artificiel. *JEP. Martigny*, Switzerland.

[3] Cutler, A., Carter, D. M., 1987. The predominance of strong initial syllables in the English vocabulary. *Computer speech and Language*, 2, 133-142.

[4] Cutler, A., Norris, D., 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113-121.

[5] Di Cristo, A., 2000. Vers une modélisation de l'accentuation du français: seconde partie, *French language Studies*, 9, 143-179.

[6] Fonogy, I., 1979. L'accent français: accent probabilitaire, *Studia Phonetica*, 15, 123-133.

[7] Mattys, S.L., Jusczyk, P.W.; Luce, P.A., Morgan,J.L, 1999. Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38(4), 465-494. 10

[8] McQueen, J.M., 1998. Segmentation of continuous speech using phonotactics. *Journal of Memory & Language,* 39, 21-46. 11

[9] Rietveld, A.C.M., 1980. French Word Boundaries, *Language and Speech*, 23(3), 289-296.

[10] Saffran, J.R., Newport, E.L., Aslin, R.N., 1996. Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606-621. 15

[11] Welby, P., 2003. The Slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation, these de Doctorat, Université Ohio.