

## **Contribution de la structure syllabique de surface à la segmentation lexicale**

**Nicolas Dumay<sup>1</sup>, Alain Content<sup>1,2</sup> & Uli H. Frauenfelder<sup>2</sup>**

<sup>1</sup> Laboratoire de Psychologie Expérimentale,  
Université Libre de Bruxelles — 50, avenue F.D. Roosevelt — CP-191 —  
B-1050 Bruxelles — BELGIQUE — e-mail: ndumay@ulb.ac.be

<sup>2</sup> Laboratoire de Psycholinguistique Expérimentale,  
Université de Genève — SUISSE

**Abstract.** This study examined the use of acoustic-phonetic cues to syllable boundaries in on-line lexical segmentation. Two word-spotting experiments manipulated the alignment between word and surface syllable boundaries. A misalignment cost was observed for obstruent-liquid clusters but not for /s/ + obstruent clusters. Durational analyses of the materials as well as of a large corpus of utterances showed systematic variations in segment durations as a function of boundary location in obstruent-liquid clusters only. We conclude that the availability of acoustic-phonetic cues to word boundary location in consonant clusters depends upon the cluster class. When available, these cues are exploited in on-line lexical segmentation of speech.

### **1. INTRODUCTION**

L'absence de marqueur spécifique et systématique des frontières lexicales dans le signal de parole amène le système de traitement du langage à exploiter de multiples indices corrélés avec les débuts de mots afin de réaliser un découpage adéquat de l'énoncé en mots. La structure syllabique du signal semble jouer un rôle déterminant dans ces processus. Ainsi, diverses études réalisées dans des langues différentes suggèrent qu'une des stratégies de segmentation utilisées consiste à postuler un début de mot à chaque nouvelle attaque de syllabe (Dumay, Banel, Frauenfelder & Content, 1998; McQueen, 1998; Norris, McQueen, Cutler & Butterfield, 1997; Vroomen & de Gelder, 1997; pour une discussion, Frauenfelder & Content, 1999).

Au regard des théories phonologiques (*cf.* Selkirk, 1982) et de certains modèles psycholinguistiques de la production de la parole (*cf.* Levelt & Wheeldon, 1994), l'heuristique de segmentation à l'attaque semble peu viable. En effet, ces théories font l'hypothèse d'un processus de resyllabation de la chaîne segmentale basé sur le principe de l'attaque maximale et opérant indépendamment des frontières lexicales (e.g. "grec ancien" > "gre.can.cien"; Pulgram, 1970). En outre, dans le modèle de Levelt & Wheeldon (1994), la parole est produite sur base de schémas syllabiques articulatoires sélectionnés après la resyllabation. Ce processus devrait donc créer en surface des décalages fréquents entre attaques de syllabes et débuts de mots, rendant l'heuristique de segmentation lexicale à l'attaque inefficace voire perturbatrice.

Contrairement à la notion de resyllabation, plusieurs études ont cependant démontré la présence d'indices acoustico-phonétiques, essentiellement de durée, quant à l'assignation de la consonne pivotale d'un côté ou de l'autre d'une frontière lexicale (Barry, 1984; Quené, 1992, 1993). Ainsi par exemple, en situation de choix forcé, des auditeurs néerlandais entendant des paires ambiguës du type "diep in" vs "die pin" (/di(ɪ)p(ɪ)n/), exploitent des variations dans la durée de la consonne pivotale et dans le

temps de montée en énergie de la voyelle qui suit la frontière, pour décider du type de segmentation entendue (Quené, 1992).

La présente recherche a examiné la nature de ces indices acoustico-phonétiques de syllabation dans divers groupes consonantiques en français, et a évalué leur rôle dans la segmentation lexicale et la reconnaissance des mots. Deux expériences de détection de mots enchâssés (e.g. *rap*) dans des paires de porteurs ambigus quant à l'assignation de la consonne pivotale – VC#CV vs V#CCV – ont été réalisées. L'expérience I a étudié les effets de l'alignement initial (e.g. *ug rap* vs *u grap*) et de l'alignement final (e.g. *rap ri* vs *ra pri*) entre mot et syllabe phonétique, sur les groupes consonantiques OBLI (cf. Dell, 1995). Les OBLI ont été comparés aux groupes '/s/+obstruante' (SOB) dans l'expérience II qui étudiait uniquement l'effet de l'alignement final.

## 2. EXPÉRIENCE I

### 2.1. Méthode

**Participants.** Trente-deux étudiants de langue maternelle française, ne présentant aucun trouble auditif ou du langage, ont participé à l'expérience.

**Matériel.** Soixante-quatre mots CVC français, substantifs ou adjectifs, étaient enchâssés comme cibles à détecter, les uns (32 mots; e.g. *tante*) à l'initiale, les autres (32 mots; e.g. *roche*) en finale de séquences **CVCCV** (e.g. *tantrou*) ou **VCCVC** (e.g. *icroche*) sans signification. Les groupes consonantiques médians des séquences porteuses étaient de type OBLI<sup>1</sup>. Deux versions acoustico-phonétiques de chaque porteur ont été obtenues, extraites d'énoncés appariés tri ou quadri-syllabiques de deux mots, différant quant à la position de la frontière lexicale: /VC#CV/ ou /V#CCV/. Dans la condition d'alignement, la cible était effectivement réalisée (e.g. "*tant rou*" extrait de *tante roublarde*; "*ic roche*" extrait de *magique roche*). Dans celle de non-alignement, la cible chevauchait les deux mots ou était enchâssée dans le second (e.g. "*tan trou*" extrait de *temps troublant*; "*i croche*" extrait de *demi-croche*). La plupart des énoncés consistaient en des groupes nominaux. Afin d'éviter tout biais lexical sur l'un des deux points de segmentation étudiés (/CC/ vs /C.C/), notre matériel était tel qu'un mot était systématiquement présent de part et d'autre de chacun de ces deux points<sup>2</sup>.

Un matériel similaire composé de 32 séquences CVCCV (e.g. "*jufran*" extrait de *juffe rembourrée* vs *jus framboisé*) et de 32 séquences VCCVC (e.g. "*ifraise*" extrait de *actif raise* vs *demi-fraise*) dans lesquelles le CVC n'était jamais un mot, a constitué les essais remplisseurs. Pour chaque position d'enchâssement, porteurs et remplisseurs ont été mélangés aléatoirement en deux listes expérimentales, chacune composée de 32 porteurs et de 32 remplisseurs, le facteur Alignement étant contre-balancé à travers les listes.

**Procédure.** Les énoncés ont été prononcés avec enchaînement des deux termes, par un locuteur de langue maternelle française. Les séquences, extraites à l'aide de Sound Designer et stockées sur un Macintosh Power PC à une fréquence d'échantillonnage de 22 kHz avec résolution 16-bit, étaient présentées toutes les 3 s. au participant. Celui-ci

<sup>1</sup> Il s'agissait des groupes *pl, pr, tr, kl, kr, bl, br, dr, gl, vl, vr, fl* et *fr*.

<sup>2</sup> Cela signifie que dans un porteur comme par exemple, "*tan(##)t(##)rou*", non seulement les parties "*temps*" et "*tante*" étaient des mots mais également les parties "*trou*" et "*rou*".

devait presser le plus rapidement possible un bouton-réponse dès qu'il détectait un mot CVC tantôt à l'initiale (une partie de l'expérience), tantôt en finale d'une séquence (une autre partie de l'expérience), et ensuite répéter intelligiblement le mot détecté.

## 2.2. Résultats et discussion

**Tableau 1.** Temps de réaction moyens (TRs) à partir de la fin de la cible, et pourcentages d'erreurs (omissions + réponses incorrectes) pour les expériences I et II. "\*\*\*" indique que l'effet est statistiquement significatif par sujets et items; "\*" indique que l'effet est statistiquement significatif par sujets seulement.

	<u>Début du mot</u>			<u>Fin du mot</u>		
	Alignement	Non-alignement	d	Alignement	Non-alignement	d
<b>EXP. 1</b>	ic <b>roche</b>	i <b>croche</b>		<b>tant</b> rou	<b>tan</b> trou	
TRs (ms)	869	967	98**	916	1056	140**
Ers (%)	13	14	1	14	16	2
	<u>OBLI</u>			<u>SOB</u>		
	Alignement	Non-alignement	d	Alignement	Non-alignement	d
<b>EXP. 2</b>	<b>tant</b> rou	<b>tan</b> trou		<b>ras</b> tu	<b>ra</b> stu	
TRs (ms)	826	924	98**	762	768	6
Ers (%)	19	23	4	18	14	4*

**Analyse des Temps de réaction.** Comme rapporté dans le haut du Tableau 1, les mots enchâssés en finale étaient en moyenne détectés 98 ms plus vite lorsqu'ils se trouvaient alignés avec l'attaque de la seconde syllabe phonétique (869 ms) que lorsqu'ils ne l'étaient pas (967 ms). De même, les mots enchâssés à l'initiale étaient en moyenne détectés 140 ms plus vite lorsque leur fin était alignée avec la fin de la première syllabe phonétique (916 ms) que lorsqu'elle ne l'était pas (1056 ms). Ces deux effets se sont révélés hautement significatifs à des ANOVAs avec l'Alignement comme facteur principal (enchâssement final:  $F(1,31) = 8.94, p < .01$ ;  $F(2,1,31) = 24.49, p < .0001$ ; enchâssement initial:  $F(1,31) = 18.15, p < .001$ ;  $F(2,1,31) = 19.07, p < .001$ ). Par ailleurs, ils ne différaient pas significativement l'un de l'autre, comme en témoigne l'absence d'interaction entre les effets des facteurs Alignement et Position ( $F(1,31) = 1.01$ ;  $F(2,1,62) = 1.26, p < .3$ ). La mise en évidence de ces coûts perceptifs à l'initiale et en finale du mot démontre la présence dans les groupes OBLI, d'indices acoustico-phonétiques de syllabation congruents avec les intentions lexicales du locuteur.

**Analyse des Taux d'erreurs.** Les Taux d'erreurs étaient sensiblement les mêmes dans les 4 conditions. Des ANOVAs similaires à celles effectuées sur les TRs n'ont révélé aucun effet ni d'interaction statistiquement significative.

## 3. EXPÉRIENCE II

L'expérience II a examiné la généralité des indices acoustico-phonétiques quant à l'assignation de la consonne pivotale, en comparant les groupes OBLI aux groupes SOB. Seul l'effet de l'alignement final a été étudié.

### 3.1. Méthode

**Participants.** Quarante-huit étudiants de langue maternelle française, ne présentant aucun trouble auditif ou du langage, ont participé à l'expérience.

**Matériel et procédure.** Soixante-quatre mots C(C)VC français (e.g. *race*), substantifs ou adjectifs, étaient enchâssés comme cibles à détecter à l'initiale de séquences **C(C)VCCV** sans signification (e.g. *rastu*), dont le groupe consonantique médian était de type SOB ou OBLI<sup>3</sup>. Comme dans l'expérience I, deux versions acoustico-phonétiques de chaque porteur ont été extraites d'énoncés tri ou quadri-syllabiques de deux mots, différant quant à la position de la frontière lexicale: soit /VC#CV/ (e.g. *race tuméfiée*) soit /V#CCV/ (e.g. *rat stupéfait*). La contrainte qu'il y ait toujours un mot de part et d'autre de chacun des deux points de segmentation (/CC/ vs /C.C/) étudiés, n'a pu être respectée dans les groupes SOB. Pour le reste, matériel et procédure étaient analogues à ceux de la condition d'enchâssement initial de l'expérience I, excepté un changement de locuteur.

### 3.2. Résultats et discussion

**Analyse des Temps de réaction.** Comme on peut le voir dans le bas du Tableau 1, seul les porteurs OBLI ont engendré un effet significatif de l'alignement (98 ms) dans la détection du mot ( $F(1,47) = 5.65, p < .025$ ;  $F(2,1,31) = 21.80, p < .0001$ ), reproduisant ainsi l'effet obtenu dans l'expérience I. Aucun effet d'alignement n'a été observé sur les porteurs SOB (6ms;  $F_s < 1$ ). La subordination des effets de l'alignement à la nature du groupe consonantique est confirmée par l'interaction significative entre les effets de ces deux facteurs ( $F(1,47) = 3.83, p < .057$ ;  $F(2,1,62) = 9.62, p < .01$ ).

**Analyse des Taux d'erreurs.** Dans le cas des porteurs OBLI, le taux d'erreurs était légèrement moins important lorsque la fin du mot était alignée avec la fin de la première syllabe phonétique (19%) que lorsqu'elle ne l'était pas (23%); cette relation s'inversait dans le cas des porteurs SOB (18% vs 14%). Cependant, des ANOVAs similaires à celles effectuées précédemment n'ont révélé aucun effet ni d'interaction statistiquement fiable.

## 4. EXPÉRIENCE III

Afin d'évaluer la nature des différences acoustico-phonétiques entre les séquences à groupe OBLI, et de confirmer l'absence de telles différences dans les groupes SOB, des analyses de durée ont été réalisées a posteriori. Sur les séquences à groupe OBLI, ces analyses ont révélé un allongement significatif de  $V_1$  (L1: 29%; L2: 20%) et de  $C_3$  (L1: 41%; L2: 47%) de manière convergente chez les deux locuteurs, ainsi qu'un allongement significatif de  $C_2$  chez le locuteur 1 (L1: 30%), lorsque la consonne pivotale était assignée en coda de la première syllabe (/V<sub>1</sub>C<sub>2</sub>#C<sub>3</sub>V<sub>2</sub>/) par rapport à la situation où celle-ci se trouvait en attaque de la seconde (/V<sub>1</sub>#C<sub>2</sub>C<sub>3</sub>V<sub>2</sub>/). Par contre, aucune différence significative n'a été observée dans les séquences à groupe SOB. L'expérience III a vérifié la constance de ces indices de durée sur l'assignation de la consonne pivotale dans les groupes OBLI, ainsi que l'absence de tels indices dans les groupes SOB, chez huit locuteurs naïfs de langue maternelle française.

### 4.1. Méthode

<sup>3</sup> Les groupes consonantiques étaient *st, sp, sk et sf* pour la classe SOB, et *tr, gr, pr, dr, br, vr, cr et fr* pour la classe OBLI.

**Matériel et procédure.** Le corpus était constitué de 48 paires d'énoncés de deux mots /CVC#CV.../ – /CV#CCV.../, dont 32 à groupes OBLI (16 'obstruante+/l/' et 16 'obstruante+/r/') utilisées dans l'expérience I, et 16 à groupes SOB utilisées dans l'expérience II. Les 96 énoncés ont été mélangés aléatoirement en 2 blocs équilibrés. La réalisation isolée des énoncés, décrits au locuteur comme des groupes de sens et non de simples paires de mots, consistait en une lecture enchaînée des deux termes. Pour la réalisation en contexte phrastique, les énoncés étaient insérés immédiatement après le verbe en tant que complément de celui-ci, dans des phrases appariées en termes du nombre de syllabes et de la position de l'énoncé. Le reste de la phrase s'accordait tantôt avec le premier terme du complément, comme dans "Aurélie est une tante roublarde du Brabant.", appariée avec "Il naquit en ces temps troublants d'avant-guerre.", tantôt avec le verbe comme dans "J'ai souffert d'une mince cataracte comme un martyr.", appariée avec "Il nettoie la main scarifiée à l'aide d'éther.". Le locuteur lisait silencieusement la phrase, puis la disait à l'expérimentateur.

#### 4.2. Résultats et discussion

**Tableau 2.** Durées segmentales moyennes en fonction du groupe consonantique, de l'assignation de la consonne pivotale et du contexte de réalisation<sup>4</sup>. Une cellule représente 128 données.

Réalisation	Type C <sub>2</sub> C <sub>3</sub>	CVC.CV					CV.CCV				
		C <sub>1</sub>	V <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	V <sub>2</sub>	C <sub>1</sub>	V <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	V <sub>2</sub>
<u>isolée</u>	Ob + /l/	106	143**	89	61**	101	105	115	83	47	99
	Ob + /r/	88	135**	63	99**	85	91	111	59	82	79
	/s/ + Ob	95	119	95	88	73	96	116	89	83	83
<u>en phrase</u>	Ob + /l/	68	122**	68	49**	97	66	102	67	41	94
	Ob + /r/	59	108**	48	87**	81	56	89	50	74	73
	/s/ + Ob	61	101	73	81	67	63	96	77	77	74

Pour les énoncés à groupe OBLI, l'allongement caractéristique de V<sub>1</sub> et de C<sub>3</sub> lorsque la frontière lexicale sépare le groupe consonantique, a été retrouvé de manière statistiquement fiable en production isolée (23% pour V<sub>1</sub>; 25% pour C<sub>3</sub>; voir Tableau 2) et en contexte de phrase (20% pour V<sub>1</sub>; 19% pour C<sub>3</sub>). Par contre, pour les énoncés à groupe SOB, aucune différence fiable en fonction de la position de la frontière lexicale n'a été observée.

#### 5. DISCUSSION GÉNÉRALE

La présence d'indices acoustico-phonétiques quant à l'assignation (VC.C vs V.CC) de la consonne pivotale dans les groupes consonantiques examinés semble dépendre de la nature du groupe. Dans les groupes SOB, les mesures de durée segmentale n'ont montré aucune différence systématique. Par contre, des variations mesurables sont présentes pour les groupes OBLI, et sont exploitées lors de la segmentation lexicale. En effet, un coût perceptif dans la détection du mot enchâssé a été induit par le non-alignement entre début de mot et début de syllabe phonétique (cf. i croche; Exp I) et par le non-alignement entre fin de mot et fin de syllabe phonétique (cf. tan trou; Exp I & Exp II). Nos résultats

<sup>4</sup> Les mesures de durée ont été réalisées par le premier auteur à l'aide du logiciel Sound Designer.

confirment donc sur les groupes consonantiques OBLI du français et au moyen d'une tâche immédiate impliquant la reconnaissance des mots, ceux obtenus précédemment dans d'autres langues au moyen de tâches de discrimination (Barry, 1984; Quené, 1992, 1993).

L'hypothèse selon laquelle les attaques de syllabes constituent des points privilégiés d'accès lexical rend compte directement de l'effet du non-alignement initial. Par un processus rétroactif, cette hypothèse explique également l'effet du non-alignement final par l'influence néfaste du mot aligné à l'attaque suivante ("trou") qui dès lors entraine en compétition avec le mot cible ("tante") pour la consonne pivotale ("t").

Les résultats obtenus pour les groupes OBLI indiquent que le phénomène de resyllabation n'a pas eu lieu. Contrairement aux hypothèses de Levelt & Wheeldon (1994), il apparaît donc que, tout au moins pour les groupes OBLI, la production d'une séquence VC#CV requiert l'activation de schémas syllabiques articulatoires différents de ceux utilisés pour produire la séquence appariée V#CCV où l'obstruante appartient lexicalement à la seconde syllabe.

## 6. REMERCIEMENTS

Cette recherche a été financée par une subvention du Ministère de la Communauté Française de Belgique (Projet A.R.C. 96/01-203) et par le FNRS-Suisse (Projet 1113-049698.96). N. Dumay bénéficie d'un mandat d'aspirant du Fonds National de la Recherche Scientifique belge.

## 7. RÉFÉRENCES

- Barry, W.J. (1984). Perception of juncture in English. In M.P.R. van den Broecke & A. Cohen (Eds.), *Proceedings of the Tenth International Congress of Phonetic Sciences, Utrecht 1983*, pp. 529-536. Dordrecht, Cinnaminson: Foris.
- Dell, F. (1995). Consonant clusters and phonological syllables in French. *Lingua*, 95, 5-26.
- Dumay, N., Banel, M.H., Frauenfelder, U.H., & Content, A. (1998). Le rôle de la syllabe: segmentation lexicale ou classification? *Actes des XXIIèmes Journées d'Étude sur la Parole*, (pp. 33-36), Martigny, Suisse.
- Frauenfelder, U.H. & Content, A. (1999). The role of the syllable in spoken word recognition: Access or segmentation? *Actes des 2<sup>èmes</sup> Journées d'Études Linguistiques*, Nantes, sous presse.
- Levelt, W.J., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50, 239-269.
- McQueen, J.M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, 39, 21-46.
- 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.
- Pulgram, E. (1970). *Syllable, word, nexus, cursus*. The Hague: Mouton.
- Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, 20, 331-350.
- Quené, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustical Society of America*, 94, 2027-2035.
- Selkirk, E.O. (1982). The syllable. In H. van der Hulst & N. Smith (Eds.). *The structure of phonological representations* (Part II). Dordrecht, Holland: Foris publications.
- Vroomen, J. & de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 710-720.