Segmentation cues in lexical identification and in lexical acquisition: Same or different ?

Odile Bagou[1, 2] *& Ulrich H. Frauenfelder*[1]

[1] Laboratoire de psycholinguistique expérimentale, University of Geneva, Switzerland
[2] Institut d'orthophonie, University of Neuchâtel, Switzerland
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 cross-modal priming paradigm, but also in the storage of new lexical representations, using an artificial language learning task. Results showed that the specific segmentation cues examined contributed differently to the performance of these two tasks. Syllable onsets, simultaneously cued by allophonic variations and phonotactics, played a predominant role in lexical identification while stress was only a "last-resort" segmentation cue. In contrast, rhythmic information, particularly primary stress, played a greater role in lexical acquisition than syllable onsets which were not used. These results suggest that segmentation cues contribute differently to these two processes and that caution is therefore required in relating results on lexical acquisition and lexical identification.

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

1. Introduction

The processes of word identification and lexical acquisition both require segmentation of the continuous speech chain. In word identification, speech must be segmented to contact and recognize the lexical representations stored in the mental lexicon. In lexical acquisition, speech must be segmented to locate the boundaries of unknown lexical units that are to be stored in the mental lexicon. The characterisation of these two processes involves determining the nature of the information sources used in segmentation that is, lexical knowledge and sublexical cues. It is quite obvious that the status of lexical knowledge differs in these two processes. In word identification, lexical knowledge is already constructed, and its contribution to segmentation has often been demonstrated [1]. Recent studies have even revealed the primacy of lexical knowledge over sub-lexical segmentation cues in word identification [6]. In lexical acquisition, the lexical knowledge of either an adult learning an artificial language [2] or of a baby [3] is initially absent and then is progressively constructed. Hence, initially segmentation is not lexically driven in lexical acquisition. An interesting question is then to establish whether the status of other segmentation cues also differs in both processes. While it is generally tacitly assumed that the use of these sublexical cues in the segmentation process in word identification and in lexical storage is very similar if not identical, this assumption must be examined directly. Our aim here is thus to determine whether the relative contribution of different sublexical cues to segmentation changes with the introduction and presence of lexical representations. We focus on the sublexical cues provided by the phonetic and phonological structure of speech.

It has been shown that in lexical identification, allophonic variation, phonotactics and rhythmic structure play an important role in identifying lexical boundaries ([4], for a review). In lexical acquisition, infants are able to segment the speech flow into lexical units [3] relying on transitional probabilities, rhythm and, later, on phonotactics. Likewise, the few studies examining segmentation with an artificial language learning paradigm suggest that transitional probabilities and rhythm are exploited by adults [5].

In most of these studies (but see [6]), each sublexical source of information has been manipulated and investigated in isolation, unlike in natural speech where multiple sources of information are simultaneously available in the speech signal. Indeed, in real-life perception, the listener continuously has to evaluate complementary or opposing cues in order to choose the segmentation which best matches the structure of the input. It is possible that a particular boundary cue is used by the listener only in controlled laboratory conditions, when other cues are absent or degraded (e.g. noisy speech [6]). We cannot be sure that the efficiency of this cue is the same when a more powerful segmentation cues is present simultaneously in the speech signal. Hence, an important objective here is to investigate how the system simultaneously integrates multiple sources of segmentation information and to evaluate the relative contribution of these multiple cues to the lexical segmentation process.

Our previous work on French has shown that syllable onsets located by both phonotactics and allophonic variation, are efficiently used in lexical identification to access the mental lexicon (SOSH, [7]). On the other hand, it has been suggested [8] that rhythmic prominence, and particularly final primary stress, function as privileged anchoring points from which the lexical units are acquired. Hence, prosodic structure appears to be used efficiently by learners of French in constructing and storing new lexical representations.

The aim of this paper is two-fold. First, it investigates how French listeners segment words when they can simultaneously make use of both syllabic alignment and prosodic boundary cues. Two results are possible: either both manipulated cues have the same weight in lexical segmentation or one of the information source is more important than the other. In the latter case, a hierarchy of these segmentation cues can be identified. Second, this paper aims at investigating the relation between segmentation in word identification and in lexical acquisition. In other words, the question is whether the efficiency of syllabic alignment

cues and prosodic word boundary cues is similar in these two processes or whether it differs depending upon the task.

Two experiments were conducted: the first examining lexical identification with a cross-modal priming task and the second investigating lexical acquisition with an artificial language learning paradigm. Two types of segmentation cues were manipulated and paired against each other in the two experiments: (1) syllabic alignment cues simultaneously marked by phonotactics and allophonic variations (S- Syllabically Aligned, NS- Non-Aligned Syllabically) and (2) prosodic-word boundary cues, examined by comparing conditions in which the last syllable before the lexical unit to identify or to acquire was bearing final primary stress or not (P- presence of primary stress or NP- No primary stress).

2. Experiment 1: Lexical identification

2.1. Method

Participants heard an auditory prime which was immediately followed by a string of letters on which participants performed a lexical decision. In the related conditions, the embedded auditory prime and the visual target were the same word, whereas they were not the same in the unrelated conditions. By comparing related and unrelated conditions, we can evaluate the amount of priming produced by the auditory prime on the visual lexical decision depending upon the manipulated segmentation cues present. We expect the amount of priming to vary according to the nature of the segmentation cues present in the auditory sequence which varied between experimental conditions (Table 1).

2.1.1. Materials

Auditory primes were created as follows: lexical and non-lexical mono- and bisyllabic words and non-words- all beginning with a liquid consonant /l/ or /ʀ/ - were inserted in nonsense carrier sequences, following the structure [σ1.word/non-word.σ2] (Table 1). To evaluate the effect of syllabic alignment cues, targets were either aligned (e.g. « lutte » in /sin.*lyt*.bad/) or non-aligned (e.g. « lutte » in /si.p***lyt***.bad/) with a syllable onset. Thus, the CVC syllables σ1 and σ2 were chosen to be aligned with the beginning of the target (S condition) or to create a cluster with the initial liquid of the target (NS condition). In addition, the effect of prosodic cues was tested by comparing conditions in which final syllables bore a final primary stress (e.g. « lutte » in **/sin**.#*lyt*.bad/) with conditions in which there was no rhythmic prominence before the word to spot (e.g. « lutte » in /sin.*lyt*.bad/). Final prominence was marked by increasing syllable's intrinsic duration by 30% and the f0 of the syllable nucleus by 40 Hz. Visual targets were letter strings corresponding to a French word or to a nonsense string respecting the orthotatic rules of the language.

Table 1 : *Experimental conditions manipulating Initial Syllabic (.) and Prosodic (bold, #) alignments.*

Priming condition	Alignment condition	Prime	Auditory Target	Visual target
Related Prime	S-P	**sin**.#lyt.bad	/lyt/	LUTTE
	S-NP	sin.lyt.bad		
	NS-P	**si**.#plyt.bad		
Unrelated Prime	S-P	**sin**.#lyt.bad	/lyt/	LAMPE
	S-NP	sin.lyt.bad		
	NS-P	**si**.#plyt.bad		

2.1.2. Procedure and task

Three experimental lists were constructed. Each participant received two of these three lists. The presentation order was counterbalanced across participants. Each prime was presented twice: once preceding a related target and once preceding an unrelated target. The proportion of related prime-target pairs was equal to 50%. Auditory primes were immediately followed by visual targets (ISI = 0). Participants were instructed to indicate as rapidly as possible by a button press whether the target letter string was a word or not (lexical decision task).

2.1.3. Participants

66 students at the University of Geneva participated in the experiment for course credits. All were native speakers of Swiss French and reported no hearing disorders. One participant with a very high errors rate (greater than 30%) was excluded from the data analysis.

2.2. Results and discussion

Analyses were performed separately for mono and bisyllabics targets. Results are presented in Table 2. Repeated measures analyses of variance (ANOVAs) were performed on lexical decision latencies measured on correct lexical identifications, by participants (F1) and by items (F2), with the "alignment condition" and the "priming condition" as within-subject factors.

Table 2: *Lexical decision latencies as a function of priming and alignment conditions (varying according to prosodic and syllabic alignment cues).*

Monosyllabics *Priming condition*	**Alignment conditions**		
	S-P	***S-NP***	***NS-P***
Related prime	539.6	535.3	526.5
Unrelated prime	518.6	521.2	532.9
Difference	**-21**	**-14,1**	**6,4**

Bisyllabics	***S-P***	***S-NP***	***NS-P***
Related prime	488.1	493.6	518.7
Unrelated prime	539.1	531.7	530.1
Difference	**51****	**38,1***	**11,4**

For monosyllabic targets, analyses revealed that neither the effect of the "priming condition" (*F1(1,64)=2.8, p=.09; F2<1*), nor the "alignment condition" (*Fs<1*) is significant. The expected effects are probably masked because of the difficulty of identifying short words in longer sequences. Indeed, we know that lexical activation depends upon the accumulation of bottom-up evidence. Hence, shorter words have a clear disadvantage over longer words [9], at least for two reasons: first, less phonemes support the activation of short words. For example, "recette" is supported by 5 phonemes while "lampe" is only supported by 3 phonemes; second, monosyllabics have more phonological neighbourhoods than longer words. For example, "lampe" is

phonetically similar to “langue”, “lente”, “rampe”) but none are equally close to “recette”. Hence, our poor results on monosyllabics are probably due to a greater lexical competition for the short words than for longer words, a suggestion that is supported by significantly poorer rates of correct lexical identifications for monosyllabics than for bisyllabics (F1(1,64)=25.5, p<.; F2(1,28)=7.8, p=.009).

For bisyllabic words, results are clear. The analyses revealed a significant priming effect (*F1(1,64)=22.9, p<.0001; F2(1,14)=4.6, p=.05*) and an interaction, only significant by participant, between the “alignment condition” and the “priming condition” (*F1(2,128)=4.9, p=.009; F2<1*). This interaction suggests that priming depends upon the nature of the segmentation information in the priming carrier sequence. More precisely, further multiple comparisons (post-hoc Tukey tests) show a significant priming effect for the aligned conditions (S-NP: syllabic alignment: p=.0004; S-P: prosodic boundary: p=.00002) while no priming is observed for the syllabically non-aligned condition (NS-P: p=.8). Moreover, for the related primes, results indicate that syllabic alignment significantly improves lexical identification (S-P vs. NS-P: p=.009) while prosodic boundary cues are not efficient in lexical identification (S-P = S-NP, p=1).

In sum, in lexical identification, syllabic alignment, simultaneously cued by phonotactics and allophonic variations, is a more powerful segmentation source of information than the rhythmic structure.

3. Experiment 2: Lexical acquisition

3.1. Method

The relative contribution of prosodic and syllabic alignment cues to speech segmentation was tested with an artificial language acquisition paradigm. Using an artificial language allows us (1) to eliminate semantic and lexical information, (2) to manipulate the prosodic properties of the “words”, (3) to control the subject's exposure to the language and, more interestingly, (4) to get adults in a situation of acquiring a language about which they have no linguistic knowledge.

3.1.1. Procedure and task

The experiment was divided into two phases: a learning and a test phase. In the learning phase, participants heard a 4 minute sequence of continuous speech consisting of concatenated artificial words with no intervening pauses. They had to extract and remember the sequences which correspond to the artificial words in the language. The participant's storage of the artificial words making up the language was tested during in a non-speeded forced-choice lexical preference task in the test phase. Participants heard pairs of either an artificial word and a non-word or of two non-words. They had to specify which member of the stimulus pair corresponds to a word of the artificial language.

3.1.2. Artificial language construction

The basic artificial language resulted from the concatenation of 16 syllables respecting French phonotactic rules (e.g. /ʃiʒ/, /syn/, /vob/, /daz/, /nul/, /bãf/, /kɛm/). Syllables were recorded by a French female speaker and re-synthesized by using the PSOLA resynthesis routine of PRAAT at 210 Hz, the mean fundamental frequency (f0) of the speaker. Eight bisyllabic artificial words beginning with a liquid were built by concatenating the 16 syllables. In addition, 8 CVC syllables, differing according to the version of the language, were added at the onset of each artificial word, resulting in the following final structure of the artificial language: [σ1.word1.σ2.word2.σ3.word3 …]

3.1.3. Experimental conditions

Four versions of the artificial language were constructed. Transitional probabilities (TP) between artificial words and additional syllables varied from .06 to .25 while the TP between syllables included in a same artificial word was 1. Moreover, as in experiment 1, syllabic and prosodic initial alignments were manipulated (Table 3). The effect of syllabic alignment was tested by adding CVC syllables which created non-aligned (NS versions) or aligned versions (S versions). In the former, the syllables were created a cluster with the initial liquid consonant of the artificial words. In addition, the effect of prosodic cues was tested by comparing cases in which final syllables were bearing a final primary stress (P versions) with conditions in which there were no rhythmic prominence before the word to be learned (NP version). Prominence was marked by increasing the syllable’s intrinsic duration by 30% and the f0 of the syllable nucleus by 20 Hz.

Table 3: *Experimental conditions manipulating Initial Syllabic (.) and Prosodic alignments (bold)*

AL Versions	Syllabic alignment	Prosodic-word boundary	Speech sample
S-P	+	+	**zɛs**.luf.**gɐd**.**kɔs**.ʁyf.**zaʃ**.**das**.lɛ̃t.**mif**
S-NP	+	∅	**zɛs.luf.gɐd.kɔs.ʁyf.zaʃ.das.lɛ̃t.mif**
NS-NP	-	∅	**ti.glyb.daz.za.klɛ̃t.mif.vɛ.kluf.gɐd**
NS-P	-	+	**ti.glyb.daz.za.klɛ̃t.mif.vɛ.kluf.gɐd**

3.1.4. Test construction

Two different tests were constructed depending on whether artificial words were aligned (S versions) or not (NS versions) with a syllable onset. Each test was made up of 80 pairs: 32 Word-NonWord pairs and 48 NonWord-NonWord pairs. Each of the eight artificial words was paired with four types of 8 NonWords sharing different syllables with the words. Only the analyses conducted on Word-NonWord pairs will be presented here, with no distinction between the NonWord types.

3.1.5. Participants

72 students at the University of Geneva participated in the experiment for course credits. All were native speakers of French and reported no hearing disorders. They were split into 4 groups of 18 participants each.

3.2. Results and discussion

Results are presented in Figure 1. Repeated measures analyses of variance (ANOVAs) were performed on mean “lexical” identification rates both, by participants (F1) and by items (F2), with “AL Version” as the between-subject factor and the “Type of pair” as repeated measures. These analyses revealed a significant effect of the AL version ($F_1(3,68)=4.4$;

p=.007; $F_2(3,21)=4.2$; p=.02), suggesting that the accuracy of the response depends upon the nature of the available segmentation cues. More precisely, further multiple comparisons (Tukey tests) indicate that prosodic cues significantly improves lexical extraction and storage (NS-NP vs. NS-P: p=.04) while syllabic alignment is not efficient in lexical acquisition (S-NP vs. NS-NP: p=1).

In sum, the rhythmic structure of the speech chain and, in particular final primary stress, is a more powerful source of segmentation information than syllable onsets indicated by both phonotactics and allophonic variation in lexical acquisition.

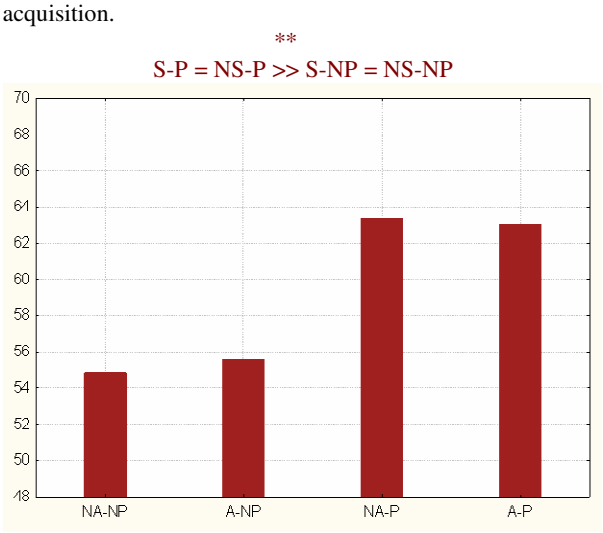

Figure 1: *% of correct lexical identification for the four AI versions of the artificial language*

4. General discussion

In this study, two experiments were conducted in order to compare the relative contribution of multiple cues to lexical segmentation and to investigate the relation between segmentation in lexical identification and in lexical acquisition. Results show that syllabic alignment, cued by allophonic variations and phonotactics, is a powerful segmentation cue in lexical identification while the rhythmic structure appears not to facilitate lexical access. In contrast, in lexical acquisition, segmentation is principally guided by the rhythmic structure, and more precisely by final primary stress, while syllabic alignment does not facilitate the acquisition of new lexical units. These results suggest that the weight of the information sources manipulated differs when the listener must consult an existing mental lexicon or construct new lexical representations. More precisely, phonotactics and allophonic variations are more powerful information sources of segmentation than stress when lexical representations are present, while rhythmic information dominates phonotactics and allophonic variations when lexical representations are not already constructed that is, in lexical acquisition. The role of the rhythmic information is particularly greater in lexical acquisition, probably because this information source is more salient and resistant to noise than the other linguistic information. Indeed, previous studies have shown that the metrical structure of the native language guides the segmentation of a foreign language while the metrical structure of the target language is not used by perfect bilinguals [10]. In contrast, non-native listeners can use their phonological knowledge of the target language in order to segment speech into words [11]. Hence, rhythmic characteristics should be more resistant to changes than phonological knowledge, thereby explaining their strengthened weight in lexical acquisition.

In conclusion, our study suggests that caution is required in relating results on lexical acquisition to those on lexical identification. Since the relative contribution of sublexical cues varies according to the task, we must be careful not to draw strong inferences about acquisition from data on lexical identification and vice-versa. It would be interesting to investigate whether and how lexical knowledge, built up on the basis of phonological and prosodic cues, relegates these initially efficient segmentation cues into the lower position of being optional cues. In the course of language development, the hierarchy of segmentation cues is indeed probably modified as lexical knowledge progressively increases.

5. References

[1] Davis, M. H. (2000). Lexical segmentation in spoken word recognition. Unpublished PhD, University of London, London.

[2] Saffran, J. R., Newport, E. L., & Aslin, R. N. "Word segmentation: The role of distributional cues", Journal of Memory and Language, vol. 35, 1996, 606-621.

[3] Saffran, J. R., Aslin, R. N., & Newport, E. L. "Statistical learning by 8-month-old infants". Science, vol. 274, 1996, 1926-1928.

[4] 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, vol. 28, 2002, 218-244.

[5] Banel, M. H., Frauenfelder, U. H. & Perruchet, P. « Contribution des indices métriques à l'apprentissage d'un langage artificiel », JEP98, 29-32, 1998.

[6] Mattys, S. L., White, L., & Melhorn, J. F. "Integration of multiple speech segmentation cues: A hierarchical framework", Journal of Experimental Psychology: General, vol.134, 2005, 477-500.

[7] Content, A., Kearns, R. K. & Frauenfelder, U.H. "Boundaries versus onsets in syllabic segmentation", Journal of Memory and Language, vol. 45, 2001, 177-199.

[8] Bagou, O., Frauenfelder, U.H. « Stratégie de segmentation prosodique: rôle des proéminences initiales et finales dans l'acquisition d'une langue artificielle », JEP06, 2006.

[9] Frauenfelder, U.H. & Peeters, G. On lexical segmentation in TRACE: An exercise in simulation, In G. Altmann (Ed.) Cognitive Models of Speech Processing: Psycholinguistic and computational perspectives", MIT Press, 1990.

[10] Cutler, A., Mehler, J., Norris, D., & Segui, J. "The monolingual nature of speech segmentation by bilinguals", Cognitive Psychology, vol. 24(3), 1992, 381-410.

[11] Weber, A. "The role of phonotactics in the segmentation of native and non-native continuous speech", Proceedings of the Workshop on Spoken Access Processes, 143-146, 2000.