Contribution of Prosody to the Segmentation and Storage of "Words" in the Acquisition of a New Mini-Language

Odile Bagou, Cécile Fougeron & Uli H. Frauenfelder

Laboratoire de Psycholinguistique Expérimentale, Université de Genève
{o.b; c.f; U.F}@pse.unige.ch

Abstract

This paper evaluates the relative contribution of two prosodic cues, lengthening and f0 contour, in the processes of speech segmentation and storage of new words. More precisely, we investigate the role of prosodic information in the acquisition by French learners of a mini-language constructed for the experiment. The results show that presence of prosodic information facilitates the speech segmentation and therefore, the acquisition of the new language. Indeed, lengthening or f0 rise on the word final syllable is used by listeners to infer the presence of a boundary. However, the presence of the two cues manipulated does not improve performance; and when only one cue is present, f0 induces slightly more accurate segmentation than lengthening. Finally, the storage of the "stressable" property of the word-final syllables in French is discussed.

1. Introduction

Prosody clearly organizes spoken language, but how humans use prosodic information in language acquisition and language processing is not yet well established. It is traditionally admitted that the acquisition of words in a new language, and the recognition of words in a known language, depend upon the segmentation of these words in the continuous speech stream (e.g. [15]).

Since no systematic acoustic cues marking words boundaries can be found in continuous speech, it remains to be explained how speech segmentation takes place, and what information types this process exploits. Two main alternatives emerge from the literature: proposals in which segmentation relies mainly on phonotactics and distributional regularities ([11]; [15]; [16]) and proposals in which segmentation depends primarily on prosodic information, more particulary for infants but also for adults (e.g. [8]; [10]).

Our goal here is to advance our understanding of how prosodic information can guide segmentation. More specifically, since we postulate that acquisition requires a prior segmentation of the speech chain, we will try do determine the contribution of different prosodic cues in the acquisition of a new language by speakers of French, a language in which prosodic information is said to be post-lexical [5, 7].

According to Cutler and Norris [4], lexical segmentation in languages with lexical stress, such as English or Dutch, is accomplished according to a Metrical Segmentation Strategy (MSS). This strategy supposes that strong syllables are used to hypothesize word boundaries. Since 90% of English words are initially stressed ([3]), using strong syllables as word onsets would be an efficient strategy.

However, to be extended to other languages, the MSS has to be adapted to operate according to the metrical properties specific to each language. Thus, we cannot apply this segmentation heuristic directly to a language like French for two main reasons. First, stress (‘*accent primaire’*) is not lexical in French: the domain of stress is a unit larger than the word, the stress group. Second, stress is not initial in French but final: it falls on the final full syllable of the last word (more often a lexical word) in the stress group. This syllable is usually marked by a final lengthening, fundamental frequency (f0) variations, and an increase in intensity [5].

Thus, any application of the MSS to French should consider: (1) stressed syllables as cues to final boundaries, rather than initial ones; and (2) the segmented units as phrases that may contain several words (2.3-2.6 words in average [e.g. 7]). Thus, simply taking stressed syllables to hypothesize word boundaries in French, would not be as efficient as in lexical-stressed languages like English. Moreover, one would need to account for the additional segmentation of the words within the stress groups.

The apparent complexity of applying MSS to French does not, however, exclude that French listeners use prosodic information to segment words in certain cases. Indeed, previous studies (e.g. [1]) have shown that French listeners exploit the presence of final lengthening to segment ambiguous one word/two words sequences (e.g. "bord#dur" vs. "bordure"). Banel et al. ([2]) also showed that final lengthening facilitates “lexical” segmentation in the acquisition of an artificial language. Indeed, French learners of a new language (constructed artificially) were more efficient in their acquisition when the language had word final lengthening rather than isochronous syllables.

The present paper is based on this work and uses the same experimental paradigm: the acquisition of an artificial language. Our main objective is to examine further the contribution of different prosodic information. Indeed, since prosodic marking is multi-parametric, several cues may contribute to segmentation. Stress ('*accent primaire*') in French is marked by final lengthening, as mentioned above, and, in sentence medial position, by a f0 rise (continuation rise). Our objective here is to investigate the relative contribution of durational and/or intonational boundary cues to the segmentation of continuous speech into "lexical" units.

Another objective is to determine whether prosody may affect the construction of the memory representation of the segmented unit. Such a process is clearly required in language acquisition for new words to be learned. In French, since stress is not lexical, it is unclear what prosodic information, if any, is stored in the lexicon, and how this information can be described.

2. Method

The contribution of prosodic cues to speech segmentation and storage was tested with an artificial "mini-language" acquisition paradigm. The use of a language constructed artificially allows us (1) to manipulate the prosodic properties of the "words", (2) to eliminate semantic and lexical information, and (3) to control the subject's exposure to this language.

Our "mini-language" (taken from [2]) consists of 8 "words" (4 bisyllabic / 4 trisyllabic) built by concatenation of 18 CVC syllables, not appearing in French but respecting its phonotactic rules (e.g. /ʃiʒsyn/; /vobdaznul/; /bɑ̃fkɛm/). The experiment was divided into two phases: a learning and a test phase.

2.1. Learning Phase

Participants heard a 12 min. continuous speech sequence consisting of concatenated "words" of the mini-language (e.g. /ʃiʒsynvobdaznulbɑ̃fkɛm.../). Each "word" appeared 100 times in the sequence in a semi-randomized order. Participants had to locate and extract words from the speech input. The participant's storage of the "words" making up the mini-language was tested as follow.

2.2. Test phase

In a non-speeded lexical discrimination task, participants had to specify which member of a stimulus pair corresponds to a "word" of the mini-language. Pairs of Word-NonWord and NonWord-NonWords, were constructed with four types of Non-Words sharing different characteristics with the words. Only the analyses conducted on Word-NonWord pairs will be presented here, with no distinction between the NonWord types. In the 80 pairs tested, each word appeared four times.

2.3. Experimental conditions

Four prosodic versions of the language (see Table 1) were constructed by re-synthesis with Praat. Version D corresponded to the French prosodic pattern of continuation: the final syllable of each word was lengthened by 30% of its intrinsic duration and carried a f0 rise (110 to 140 Hz). The relative contribution of the two prosodic cues was tested with versions B and C: version B has only final lengthening and version C, only a final f0 rise. These three prosodically specified versions were compared to a neutral version (A) constructed without final f0 rise (flat f0 contour) and final lengthening.

	Prosodically unspecified	Prosodically specified		
	A	B	C	D
Final F0 rise	-	-	+	+
Final lengthening	-	+	-	+

Table 1: Four versions of the language varying in the presence of prosodic information.

108 Swiss French participants were split into 9 groups of 12 participants each: 5 Test groups and 4 Control groups. The former were tested after being exposed to the language during the learning phase, and the latter received the test phase without a prior learning phase. Four Test groups, subgroups AA, BB, CC, and DD were tested on words containing the same prosodic information as those learned in the learning phase (e.g. AA: prosodically unspecified in both test and learning phase). These Test subgroups were matched with 4 Control subgroups on the prosodic characteristics of the words in the test phase (ØA, ØB, ØC, ØD respectively). Finally, an additional Test subgroup (DA) was used to control for a bias toward an "acoustical form" identification strategy, and to test our hypothesis on storage. In this subgroup, participants were exposed to the version of the language containing both prosodic information sources manipulated (D) and were tested on a neutral version of the words (A).

3. Results

The contribution of prosodic information to the acquisition of the mini-language was tested by comparing the performance of participants exposed to the language in a learning phase (Test groups) to that of participants not exposed to the language (Control groups). Figure 1 shows the percent of words correctly identified by Test (grey bars) and Control groups (white bars) for the 4 versions of the language (A, B, C, D). Overall the Test group performed significantly better than the Control group (+30%). This suggests that participants in the Test groups have at least partially "learned" the language. However, this increase in performance was larger for participants exposed to prosodically specified versions (B:+34%, C:+39%, D:+37%) compared to participants exposed to the neutral version (A:+13%). T-tests on recognition scores in Control vs. Test groups showed a significant effect of learning both by item and by subject in the three prosodically specified versions of the language, but only by subject in version A (see table 2a). Furthermore, half (6/12) of the participants exposed to this neutral version (A) of the language performed at chance level according to a binomial test (p<.05), while none did in version D and one in version B and C. In sum, the presence of prosodic information appears to have facilitated the acquisition of the language.

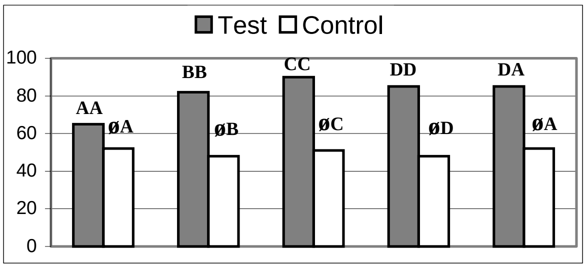

Figure 1: % correct identification in Test (grey) and Control (white) groups depending on prosodic conditions.

a. T-test: Comparison Test vs. Control Groups		
	by subjects t(22)	by item t(14)
A	3.79 (p=.001)	1.58 (p=.14)
B	7.54 (p<.0001)	7.57 (p<.0001)
C	11.16 (p<.0001)	7.54 (p<.0001)
D	10.81 (p<.0001)	8.60 (p<.0001)
b. T-test: Comparison between prosodic conditions		
	by subjects t(22)	by item t(14)
BB vs. CC	- 2.21(p=.04)	- 1.71 (p=.11)
BB vs. DD	-0.84 (p=.41)	-0.79 (p=.44)
CC vs. DD	1.18 (p=.25)	1.01 (p=.33)
DD vs. DA	0.14 (p=.89)	0.15 (p=.89)

Table 2: Summary of statistical results

The relative contribution of the different prosodic cues to language acquisition is shown by the comparison of the performance in version B, C and D (Fig. 1, grey bars). Surprisingly, better scores were found for participants exposed to a language with only final f0 rise (C: 90%) than with only final lengthening (B: 82%). However this difference was only significant by subject when the scores are compared between prosodic conditions (see table 2b). Interestingly, performances in the version in which both intonational and durational cues are present were not significantly higher (D: 85%), suggesting that the use of each type of prosodic cues is not cumulative.

The final analysis aims at testing whether learners did actually construct an abstract mental representation of the words of the artificial language or whether they only stored “acoustical forms”. In the latter case, participants in the test phase would only recognize forms that are acoustically identical (i.e. prosodically similar) to the one stored (condition DD). In the former case, participants should be able to recognize words learned with some prosodic information even if this information is not present in the test stimuli (condition DA). A comparison of the performance in the subgroups DA vs. DD shows similar identification rates (85% for both). Thus, words learned in a version of the language with prosodic information (D) appear to be learned equally well whether tested with or without these prosodic cues.

4. Discussion and conclusion

Results show that prosodic information facilitates participants’ acquisition of a new mini-language. In version A, participants could only rely on phonotactics and distributional characteristics of the language to isolate the words presented in the artificial language. Their poor performance in the test phase shows that this information was not sufficient on its own to segment the speech chain. In contrast, the presence of prosodic information located on the final syllable of the words, in version B, C, and D, appears to have facilitated the acquisition of the language. Indeed, participants performed relatively well in the test phase.

Since prosodic information is multi-parametric in French, it is interesting to weight the relative contribution of the different prosodic cues in this process. Participants’ performance in subgroup BB are comparable to that of Banel et al. ([2]) who found 85% correct identification when the artificial language followed the iambic French pattern (short-short-long). Therefore the effect of final lengthening on segmentation was replicated. However, an unexpected result was that when only one cue was present (f0 rise or final lengthening), the intonational cue induces slightly more accurate responses. Even if this effect is only significant by subject, it goes against Rietveld [14] who found that lengthening was a better cue to the segmentation of ambiguous French sentences than f0. One possible interpretation of our finding relates to the fact that the f0 rises at the end of each word in the speech sequence presented in the learning phase corresponded to the intonational contour of a list. The recurrence of these continuation rises every 2 or 3 syllables (recall that words are bi- or tri-syllabic) may have favored the lexical segmentation process. Indeed, the listeners may have hypothesized that they were listening to a list of items rather than sentences. Thus, the real weigth of the f0 variations has to be investigated further in other prosodic contexts which do not favour one segmentation over another.

The main question addressed in our study is how prosody has contributed to the acquisition of the artificial language. We assume that when learners are presented with a new language in a continuous sequence, they first have to segment the speech stream to isolate units, and then construct a mental representation of these units in their lexicon. Regarding the segmentation process, prosodic information can guide listener’s attention to specific part of the signal such as the word-final salient syllables of our mini-language ([13]). Therefore, the knowing of the position of stress can have helped listener to divide continuous speech into separate words. Indeed, even if final stress is not lexical in French, French listeners have been shown here to apply a segmentation strategy based on prosodic cues, such as MSS, in the acquisition of a new language. This could suggest that prosody is exploited pre-lexically and is used bottom-up, while lexical information should be constructed from the signal via contact with a higher level of representation. However, further studies are needed to evaluate the contribution of intonational and durational variations in natural language when phrase boundaries do not correspond to word boundaries.

Regarding the construction of the lexicon, it is doubtful that final f0 rise and lengthening as such are stored in lexical entries in French. Even if final syllables of lexical words are known to be "stressable" ([6]), we have already said that stress assignment in French is post lexical. The results we found by comparing performance of the DD and DA subgroups have shown that learners of a language containing prosodic information, are able to recognize the learned words when presented with a different prosodic contour. It is thus probable that the prosodic specifications of these words (here final f0 rise and final lengthening) were not stored in the lexicon, and prosody was only used to segment the chain. However, since several studies have shown that performance was better when the phonological properties of the artificial language matched those of the native one ([16]), we can hypothesize that some prosodic information is present in the mental representations. For e.g., we can not reject the possibility that the mental representation of the words contains some prosodic features. These features could be unspecified, allowing the matching of various surface forms with the underlying representation. This discussion raises the question of what is meant by "stressable" syllables in French words, and how this information is represented in the lexicon.

5. References

[1] Banel, M. H.; Bacri, N., 1994. On metrical patterns and lexical parsing in French. *Speech Communication*, 15, 115-126.

[2] Banel, M. H.; Frauenfelder, U. H.; Perruchet, P,. 1998. Contribution des indices métriques à l'apprentissage d'un langage artificiel. *XXIIème Journées d'Etudes sur la Parole. 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., 1999. Intonation in French In Hirst, D. and DiCirsto, A. Eds., *Intonation systems: a survey of twenty languages*. Cambridge, Cambridge University Press.

[6] Garde, P., 1968. *L'accent*. Paris, PUF.

[7] Jun, S-A; Fougeron, C., 2000. A phonological model of French intonation. In Botinis, A. ed. *Intonation: Analysis, Modeling and technology,* 209-242. Dordrecht, Kluwer Academic Publ.

[8] Jusczyk, P. W., 1997. *The discovery of spoken language*. Cambridge, MIT Press.

[9] Lehiste, I., 1972. The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51(6), 2018-2024.

[10] 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.

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

[12] Nakatani, L. H.; Dukes, K. D., 1977. Locus of Segmental Cues to Word Juncture. *Journal of Acoustical Society of America*, 62, 714-719.

[13] Pitt, M.; Samuel, A., 1990. The use of rhythm in attending to speech, *Journal of Experimental Psychology: HPP*, 16(3), 564-573.

[14] Rietveld, A.C., 1980. Word boundaries in the French language. *Language and Speech*, 23(3), 289-296.

[15] 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.

[16] Vroomen, J.; Tuomainen, J.; deGelder, B., 1998. The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, 38(2), 133-149.

6. Aknowledgments

This research was supported by a grant from the Fonds National Suisse de la Recherche Scientifique (FNRS 1114-059532.99). We especially thank Marie Hélène Banel for sharing her corpus and all the participants tested.