Only Prosody? Perception of speech segmentation in Kabyle and Hebrew

Amina Mettouchi[1], Anne Lacheret-Dujour[2], Vered Silber-Varod[4], Shlomo Izre'el[3]

1: Laboratoire de Linguistique de Nantes, Université de Nantes, et Institut Universitaire de France <amina.mettouchi@univ-nantes.fr>;
2: Laboratoire MoDyCo, Université de Paris X-Nanterre, et Institut Universitaire de France <anne@lacheret.com> ;
3 & 4: Department of Hebrew Culture Studies, Section of Semitic Linguistics, Tel Aviv University <izreel@post.tau.ac.il>; <vereds@openu.ac.il>

Abstract

The aim of the present study is to assess the importance of prosody in the perceptual delimitation of "units" for the spoken language, by resorting to experiments involving non-speakers of the language, filtered speech, and automatic segmentation by the software ANALOR. The three experiments test in various ways the lack of access to semantic and syntactic information, as opposed to the expert's segmentation. Results show that quantitatively stronger prosodic cues are needed for informants without access to the syntax-semantics of the sample, especially when they are non-speakers. The analysis also suggests the existence of both universal and language-specific prosodic cues.

1. Introduction

One of the fundamental questions underlying the analysis of spoken languages is their decomposition into units that can be considered basic in terms of informational processing and communication. Even once the importance of prosody has been recognized, the basis of the decomposition remains problematic, because of the number of cues that play a role in the segmentation of the flow of speech into units.

The aim of the present study is therefore to assess the respective import of prosodic and non-prosodic cues in the perceptual delimitation of "units" for the spoken language, by resorting to

experiments involving non-speakers of the language, filtered speech, and automatic segmentation.

Our study deals with two geographically separated but genetically-related (Afroasiatic) languages: Kabyle (Berber, spoken primarily in Kabylie, Northern Algeria) and Hebrew (Semitic, spoken primarily in Israel).

2. Experimental Procedure

Two samples[1] were extracted from spoken narratives, one conversational (Hebrew, 30.65 sec.), the other a traditional tale (Kabyle, 32.78 sec). Those samples were segmented into intonation units (henceforth: IU), on the basis of perception, by speakers of the language. The number of perceived IUs was 21 for Kabyle and 25 for Hebrew.

The first experiment ("A") is a non-speaker segmentation of the recording, the second one ("B") is a native speaker segmentation of a filtered version of the same recording, and the third one ("C") is an automatic segmentation using the software ANALOR. The three methods enable us to examine perceptual segmentation using prosodic cues only, without access to the semantic, syntactic, informational and pragmatic contents of the speech sample, as opposed to the expert segmentation.

2.1. Non-speaker segmentation (A)

Experiment A took place during a summer school in Corpus Linguistics held in Nantes in June 2006 (<http://www.letters.univ-nantes.fr/lling/elco/>).

44 informants took part in the experiment. Among them were native speakers of French (the majority) and other European and non-European languages. The participants were linguists or students of linguistics, among whom more than half (27) had at least some experience in transcription, yet not necessarily in IU segmentation.

In experiment A the informants were asked to listen and mark IU boundaries[2] according to their perception, on a transcription divided

[1] For glossed and translated versions of the respective samples see http://clf.unige.ch.

[2] They were asked to mark "intonation groups" ("groupes intonatifs") by introducing slashes on a printed text according to the recording they heard.

into words according to the standard orthography of the respective languages (which, in the case of Hebrew, where the orthography is not based on Roman characters, was introduced unto the phonetic transcription).

The hypothesis underlying experiment A is that non-speakers' segmentation will rely on purely prosodic cues, and will therefore occur only where prosodic cues are present.

2.2. Filtered speech segmentation (B)

Experiment B was held separately at Tel-Aviv University and at the Centre de Recherche Berbère (INALCO, Paris). The Tel-Aviv experiment was conducted during a graduate seminar. 12 students, all of them fluent speakers of Hebrew, took part in the experiment. 10 were native speakers of Hebrew, 2 had Arabic as their native tongue. Only one of the informants had any previous experience in IU segmentation, and only a few of them had any experience in transcribing at all.

The Paris experiment was conducted before a seminar on spoken corpora. 6 native speakers of Kabyle took part in the experiment. 3 were members of the staff (2 professors and one librarian), 3 were students (Master's Degree in Berber). All of the 6 informants were Kabyle-French bilinguals, 2 knew Arabic as well. None had previous experience in IU segmentation, and only one had experience in transcription.

A filtered sound was achieved by "pass Hann band" process, one of PRAAT[3] features that converts every selected sound object into a filtered sound. The filtering values were from 300Hz to 600Hz (with a standard value of 100Hz smoothing). Those values were chosen by perception, i.e. at those pass band values the sound segments were unintelligible and informants were only able to recognize the syllabic pattern of the sound (this technique was already implemented for Hebrew in Silber-Varod (2005)). The filtered sample of Kabyle was obtained by the same method. The filtering values were from 300 to 500 Hz (with a standard value of 100Hz smoothing).

Experiment B shares with experiment A the hypothesis that the informants will rely solely on prosodic cues for segmentation. But B-

[3] Paul Boersma and David Weenink, www.fon.hum.uva.nl/praat/

informants differ from A-informants in that they are listening to the prosodic cues of their own language. If prosodic cues are to a certain extent language-specific, the B experiment should underline it.

2.3. Automatic segmentation (C)

Automatic segmentation of both recordings was conducted with the software ANALOR (Lacheret & Victorri 2002; Avanzi, Lacheret & Victorri 2007), first developed for the prosodic analysis of French. The method of analysis relies on the acoustic segmentation of the melodic line and the analysis of pause duration.

Practically speaking, ANALOR allows the recognition of two types of salience:

a) Strong instances of salience (converging prosodic parameters), considered as marking the end of units ('periods') signaling the way in which the speaker organizes the conceptual or communicative packaging of the message.

b) Less remarkable points of salience associated with the identification of prominences in terminal syllables (isolated cues: F0 variations, or syllable or pause duration).

Experiment C has two functions: by providing precise values for acoustic parameters at supposed IU boundaries, it allows us to compare perceptual segmentation to actual acoustic prominences; also, by proposing possible boundary points, ANALOR is an 'informant' which has no access to other than prosodic material, and whose parameterization is transparent.

3. Results

Tables 1 and 2 summarize the results in Kabyle and Hebrew respectively. Line A summarizes the results of the non-speakers segmentation (Experiment A; n=44); line B summarizes the results of the filtered speech (Experiment B; n=6 in kabyle and 12 in Hebrew). Figures in rows A and B of the tables correspond to the percentage of alignment between the informants' segmentation, and the expert's one. All the results for human informants are expressed in percentages for the sake of comparability, even if the small number of informants that took part in the Kabyle B experiment precludes strong generalizations. Line C indicates the result of the ANALOR segmentation (Experiment C), with the parameters involved in the

recognition of a boundary. 1/0 means that a boundary is/isn't detected. For detected boundaries, strong instances of salience were coded (P) or <P>. Less remarkable instances of salience were referred to by their cue.[4] Pauses are indicated by their length (in ms).

	IU1	IU2	IU3	IU4	IU5	IU6
A	7	86	66	36	82	54
B	0	83	66	0	100	16
C	1, F0	1, <P> +D,+G, =S	1, <P> +D, ++G, ++S	1, D	1, (P) +D, =G, -S	1, (P) -D, +G, +S
Pause	-	395 ms	403 ms	-	319 ms	-

IU7	IU8	IU9	IU10	IU11	IU12	IU13	IU14
66	9	88	26	52	21	61	52
16	0	100	0	100	16	83	16
1, F0	1, F0	1, <P> +D, +G, ++S	1, F0	1, <P> =D, +G, +S	1, F0	1, (P) +D	1, D
90 ms		367 ms		1060 ms		611 ms	451 ms

IU15	IU16	IU17	IU18	IU19	IU20	IU21
16	81	88	66	69	33	end
0	83	66	33	33	0	end
1, F0	1, <P> +D, +G, +S	1, <P> ++D, +G, =S	1, <P> +D, +G, ++S	1, (P) =D, +G, =S	1, <P> =D, +G, +S	end
		781 ms	458 ms			

Table 1: results of 3 experiments (A, B and C) on Kabyle

[4] Abbreviations are: (P) probable period; <P> definite period; F0 fundamental frequency; D syllable or pause duration; S pitch reset; G amplitude of the final melodic gesture compared to mean F0 values inside the period. A default detection threshold is set for the various parameters. If the threshold is attained, this is marked by an 'equal' sign. If the values are higher than the threshold, we indicate this by a '+' sign, and if they are much higher, by a '++' sign.

	IU1	IU2	IU3	IU4	IU5
A	39	7	64	48	82
B	100	58	83	92	100
C	1, <P> ++D,+G, ++S	0 D	1 <P> -D, ++G, S	1 (P) -D, ++G, =S	1 <P> +D, +G, +S
Pause	1139	-	355	659	364

IU6	IU7	IU8	IU9	IU10	IU11	IU12
2	86	48	23	2	93	80
42	100	67	42	42	100	92
0 F0	1 <P> +D, ++G, +S	1 F0	0 Fall, IG	0	1 <P> +D, +G, ++S	1 (P) +D, ++G, -S
-	473	-	-	-	427	146

IU13	IU14	IU15	IU16	IU17	IU18
93	18	11	7	86	66
100	25	33	25	100	100
1 (P) ++D, ++G, -S	0	0 Fall, IG	1 F0	1 <P> +D, ++G, ++S	1 <P> +D, +G, =S
1572	-	-		949	434

IU19	IU20	IU21	IU22	IU23	IU24	IU25
-	-	-	-	-	-	end
67	67	25	8	42	8	end
1 F	1 F0	1 F0	1 F0, D	1 F0	0 Fall, IG	end
						end

Table 2: results of 3 experiments (A, B and C) on Hebrew

Figure 1 is the graphic illustrations of this comparison between the three experiments in Kabyle and Hebrew.

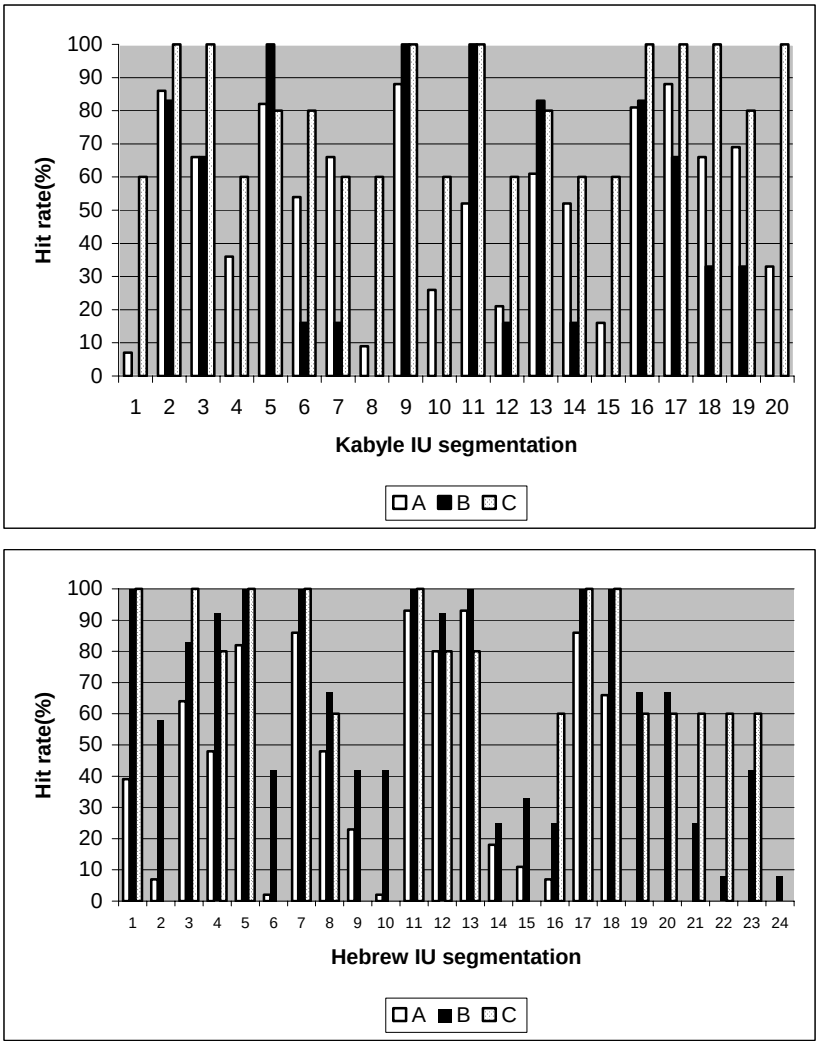

Figure 1: Comparison of hit rate according to expert's among the three experiments. In order to differentiate between strong and weaker prominences, Analor detection is marked arbitrarily as <P>= 100%, (P)= 80% and other detections= 60%.

Table 3 summarizes the degree of convergence between informants and expert in the two languages. We consider here 50% as the limit between significant and less significant perception of a boundary set by the expert.

Informants "hit" cases	*Kabyle (20 IUs)*	*Hebrew (18 IUs)*
Case 1: A & B > 50%	IU2, IU3, IU5, IU9, IU11, IU13, IU16, IU17	IU3, IU5, IU7, IU11, IU12, IU13, IU17, IU18
Case 2: A & B <50%	IU1, IU4, IU8, IU10, IU12, IU15, IU20	IU6, IU9, IU10, IU14, IU15, IU16
Case 3: A < 50% but B >50%	-	IU1, IU2, IU4, IU8
Case 4: A > 50% but B < 50%	IU6, IU7, IU14, IU18, IU19	-

Table 3: Convergence in IU detection among informants and expert

3.1. Kabyle

Hit rate in the two perceptual experiments is relatively high (over 50% hits): 12 out of 20 IUs (60%) in A; and 8 out of 20 IUs (40%) in B.

3.1.1. ANALOR and the expert's segmentation

ANALOR detected all the IUs considered as such by the expert, although all boundaries were not considered as having equal status: 8 IUs were calculated as <P>, 4 IUs as (P), and the remaining 8 IUs are minor boundaries involving F0 or duration. Those results show that the expert's segmentation was based on prosodic cues, though not necessarily only on them, since some boundaries were deemed minor by Analor, and major by the expert. Does this mean that the expert's segmentation also relied on non-prosodic cues? Or that it relied on isolated prosodic cues that might be language-dependent (F0 direction, lengthening, intensity)? Or both?

3.1.2. Expert vs. Non-speakers segmentation

Table 3 summarizes the degrees of convergence/divergence. Although all the IU boundaries were recognized by at least 7% of informants in group A, good convergence between perception by informants and expert detection is observable for only 40% of the IUs for Kabyle (case 1 in Table 3). The IUs where convergence between informants and expert is strong are also those that are detected by ANALOR as having a strong or rather strong prominence (<P> or

(P)). This shows that the more numerous the cues are, the better for informants relying only on prosodic cues (quantity principle).

For some IUs, the informants' segmentation did not meet the expert's segmentation. In this case, either both groups (non-speakers and speakers) failed to recognize a boundary (case 2 in Table 3), or only one group failed in the task (cases 3 and 4 in table 3). When both groups failed (case 2) it is not because of lack of prosodic information, since we can notice that ANALOR detected a prosodic boundary, even if it was a minor one (isolated cues: F0 or syllable/pause duration). An exception to that tendency is IU20, which is prominent (detected as a period <P> by ANALOR, and marked by the expert as a strong boundary), and was detected only by 33% of group A, and no one in group B. IU20 and IU19 are identical syntactically, but they differ as follows: IU19 has a longer final syllable, at 230Hz and 70dB; IU20 has a medium length final syllable, at 274 Hz and 63 dB. We suggest that A and B were more sensitive to lengthening, intensity and/or pauses than to F0.

When only one group failed, it was always B (case 4). No pattern can be found here. We suggest that lack of pattern is due to the statistical unreliability of the group (only 6 informants). Tendencies could be drastically different for a larger group.

3.1.3. The role of pauses

The length of pauses in our sample goes from 90 ms to 1060 ms. The mean length between those extremes is 426 ms, and most of the pauses are between 350 and 450 ms. The expert's segmentation is based both on pauses and other cues, while pauses were very significant for groups A and B. Looking more precisely at the results obtained by B, we can see that they systematically failed (0 or one informant only marked a boundary) when the cues were only either F0 or syllable duration, and systematically succeeded when there was a pause after the IU, or as in IU13 when F0 and intensity decreased significantly. However, pauses are not necessarily decisive (although they play an important role) in the detection of IUs by group A: IU7, which was not followed by a pause, was well detected by group A (66%). Pauses therefore seem to be important cues for segmentation, especially when there is limited access to F0 (filtered speech) and when there is no access to syntax and semantics (A and B).

3.2. Hebrew

Hit rate is relatively high (over 50% hits): 8 out of 18 IUs[5] (44%) among non-speakers in A; and 14 out of 24 IUs (58%) in B.

3.2.1. ANALOR and the expert's segmentation

ANALOR detected 17 out of 24 IU boundaries (71%) of which 8 are of Case 1 (see Table 3), i.e. segmented by informants of both A and B experiments; 3 were segmented by speakers but less than 50% of non speakers (case 3); 1 (IU16) was segmented <50% in both A and B experiments (case 2) and a sequence of the remaining five IUs (IU19, IU20, IU21, IU22 and IU23) were detected by ANALOR but were not part of A experiment (see Table 3).

The remaining 29% (IU9, IU10, IU14, UI15, and IU6 and IU 24 to a lesser extent) that ANALOR has not detected can be explained either due to syntactic-semantic information or due to internal prosodic cues (e.g. isotony, pitch accent). As to IU2, experiment B might prove that indeed there have been prosodic cues at this boundary[6] (see Table 2): IU2 shows a high hit percentage in experiment B, with ANALOR detecting a durational parameter. In addition, A's informants marked significantly fewer boundaries after IU2 comparing to the perception of the speakers. Admittedly, although the last syllable [Tv] shows significant length, one should note that the only major cue present in this IU is duration, this boundary was suspect also for the experts, as explained in Amir, Silber-Varod and Izre'el (2004, 678).

3.2.2. Speakers vs. non-speakers segmentation

Table 3 reveals striking results concerning speaker vs. non speaker segmentation of Hebrew. As mentioned, Speakers segmented 66% of expert's IUs (8 in case 1, 4 in case 3 and IU19, IU20). Non speakers segmented only 44% of the experts's IUs, all of them in case 1, and with very strong prosodic cues, i.e. periods. When speakers failed to "hit" expert's segmentation (case 2), it is explained because of lack of prosodic cues (see previous paragraph).

---

[5] Experiment A was taken until the last word, *gitn*, of IU18.

[6] The major acoustic cues found at IU boundaries in studies of spoken Hebrew are: (1) final length; (2) initial rush; (3) pitch reset; (4) pause (Laufer 1987; Amir, Silber-Varod and Izre'el 2004).

By far, speakers have come up with a larger percentage of hits and with fewer markings in mid-IU positions than non speakers, which seems to be a significant find. Admittedly, the conditions have not been the same, and it would be very interesting to conduct a filtered speech experiment with non speakers as well.

3.2.3. The role of pauses in Hebrew

Absence of pause is the most prominent correlate for low percentage of hits in experiment A. In these cases, at least one other acoustic boundary cue was present. For example, IU8 presents – apart from a high pitch reset [-89 Hz] – also lengthening of the last syllable and a repetitive negation in fast speech following in IU9. These two parameters, lengthening and fast speech, were not detected by ANALOR and can explain the prosodic segmentation in those IUs.

As for experiment B, among 14 IUs with no following pause, 12 present 25% of hits or more, out of which 5 present more than 50% hits. The spread of acoustic cues among these IUs is not even, and no tendency of cue hierarchy or prominence can explain the different rate of hits.

3. Conclusions

While the experiments reported here were conducted on small speech samples, we can still suggest some working hypotheses for future research along similar lines:

1. The relatively high rate of hits in all three experiments allows us to suggest that prosody is a major feature of language for segmentation perception, regardless of the language.

2. Non speakers have less access to the prosodic structure of a language than speakers, they tend to react more strongly when several prosodic parameters occur together to indicate a boundary. This can be accounted for by differences in prosodic structure among languages.

3. Experts rely on more varied prosodic cues for their segmentation. They are sensitive to F0 variations as well as to other cues, including syntax and semantics.

4. One should look further for other acoustic cues, probably less prominent than the major ones used for studying the segmentation in these experiments.

References

AMIR, N., V. SILBER-VAROD, & IZRE'EL, S. (2004), « Characteristics of Intonation Unit Boundaries in Spontaneous Spoken Hebrew – Perception and Acoustic Correlates », in B. Bel & I. Marlien (eds.), *Speech Prosody 2004, Nara, Japan, March 23-26, 2004: Proceedings*, SProSIG (ISCA Special Interest Group on Speech Prosody), 677-680.

AVANZI M., A. LACHERET & B. VICTORRI (2007), « Analor, un outil d'aide pour la modélisation de l'interface prosodie-grammaire », Actes du colloque CERLICO, Nantes, à paraître.

LACHERET A. & B. VICTORRI (2002), « La période intonative comme unité d'analyse pour l'étude du français parlé : modélisation prosodique et enjeux linguistiques », *Verbum, XXIV,* 55-72.

LAUFER A. (1987), *Intonation*. Jerusalem, the Hebrew University of Jerusalem, the Institute of Jewish Studies, Hebrew Language Department. (Hebrew)

SILBER-VAROD, V. (2005), *Characteristics of Prosodic Unit Boundaries in Spontaneous Spoken Hebrew: Perceptual and Acoustic Analysis*, MA thesis, Tel-Aviv University. (Hebrew; English abstract)