Acoustical cues as boundary markers in a left-headed language

Valéria Krepsz¹, Anna Huszár²

¹Humboldt-Universität zu Berlin, ²Hungarian Research Centre for Linguistics,
valeria-krepsz@hu-berlin.de, huszar.anna@nytud.hu

Abstract

Some previous studies have investigated the relationship between syntactic and prosodic structure, especially in those structures that are structured differently. One of the most commonly used methodologies is the variously grouped linguistic units (e.g. name grouping), which has been applied to many languages. The present examination is the adaptation of Wagner’s analysis [2] from English to Hungarian. The aim of the present study is to examine how boundaries of different strengths realize in Hungarian based on the variously coordinated names. The main question of the examination is how the prosodic boundary marker patterns differ from each other regarding their position and syntactic structure.

Duration, mean fundamental frequency of the names, appearance of silent pause, mean intensity and the appearance of creaky voice were measured regarding the names’ structure (540 items based on 15 speakers reading).

The results showed that more boundary markers appeared if the given structure contained a single boundary and/or it was a left-headed structure. Based on the examination of the similarities and differences between the languages, the results of the research contribute to the teaching of Hungarian as a foreign language and to the more accurate operation of the text-to-speech process.

Keywords: name grouping, boundary cues, prosody, acoustics, syntax-prosody interface

1. Introduction

In the past decades, there has been a growing interest regarding the interplay between the syntactic and prosodic structure of a given utterance. Previous studies tried to seek an answer as to how differently coordinated or embedded language units are realized based on the prosody of a given language. These research were conducted on some languages: for example in English Ladd [1] and Wagner [2], Féry and Truckenbrodt [3], Huttenlauch et al. [4] for German, Féry and Kentner [5, 6] for German and Hindi. Among the first researchers, Wagner [2] introduced a methodology, where the speech units either appeared as equals or implemented differently built-in categories and groups at different depths. Reading aloud these structures forces speakers to label the systematically varying boundaries after the separated speech units, as well as to express a sense of belonging names by using divergent prosodic boundary cues.

According to the latest model of speech boundaries, the interconnection of the speech units or oppositely, the signaling of the distance between two linguistic elements are due to the principles of “Proximity” and “Similarity” (for details see [5]). The concept of Proximity is based on the adjacent grouped elements’ syntactic constitution and it reflects the syntactic boundaries regarding among others the pitch and duration in the prosodic structure of a given linguistic unit. In contrast, Anti-proximity means that elements in a separate group are realized with prosodic distance, by implementing boundary markers between the groups. In the case of the Similarity concept, syntactic embedding is determined by the similarity of the linked elements: the speech units that belong to a group are similar to each other, while the elements that belong to another group are different from them (and similar to each other; Anti-Similarity).

Results from previous studies have often been contradictory as to which prosodic features occurred the most systematically and which parameters were the most relevant for boundary marking. For German [6] lower pitch and shorter word durations were observed between contiguous elements, while at boundaries, an increase in f0 and the duration of the last word was observed. Other research [7] using the Kiel corpus of German speakers found that marking of phrase boundaries occurred most frequently by the alteration of the f0 (74.0%), followed by the phrase final lengthening (66.2%), and the least frequently by pausing (38.3%). The authors conclude that, although pauses may have an important role in the detection of boundaries, its appearance is not mandatory. In French, the intonation phrase is marked by a right-edge tone that is connected to the phrase-final lengthening, while the appearance of silent pauses were also optional [8, 9].

Żygis and colleagues [10] also studied prosodic boundary marking at different normal and fastened
articulation rates. Their results showed that the alteration of the given speech units’ duration was the most dominant boundary marker, i.e. the duration of the word before the boundary increased prior to boundary marking for both speech rates. In addition to the duration change, the appearance of silent pauses also had a near-systematic boundary marking function, but its effect was less dominant for increased speech rates. In addition, intensity was also found to be a determinant factor: the intensity decreased right before the boundaries, but the effect was less pronounced for faster speech. Finally, the change in the fundamental frequency values was found to be negligible, i.e. it did not play a decisive role in the boundary signaling.

Petrone and colleagues [9] analyzed the occurrence of silent pauses, phrase-final lengthening and the f0 as prosodic boundary markers in speech production and perception. Their results showed that while the occurrence of silent pauses is mainly used to indicate intonational phrases, pre-boundary lengthening and changes in f0 values are used for other boundary marking, as well. Based on the perception test, all three prosodic boundary had an influence on listeners’ judgements: the effect of pause duration was found to be the most systematic, with variation in this marker showing categorically different responses from listeners on the coherence of the linguistic element, while there were smaller differences in the judgments regarding the f0 variation and the phrase final lengthening.

In summary, all previous studies have corroborated the role of silent pauses, longer duration (phrase-final lengthening), and change of f0 as prosodic boundary markers. However, there are notable differences as to which parameter is considered to be the most dominant or the most systematically represented in terms of either speech production or perception. The differences may be due to the characteristics of the given language, the methodology used, or partly to other factors.

The interaction between syntactic structure and the planning of prosodic boundary markers has been investigated mostly for English and German, but not for Hungarian. However, many prosodic features of Hungarian differ from the two languages mentioned: Hungarian prosodic system is relatively rich that can indicate and parse the syntactic structure of utterances (e.g. [10]). It is a topic-prominent language, the word order is mostly determined by the information structure, and defined by pragmatic factors, while prosodic prominence marking plays only a secondary role and is partly optional [11].

In addition, Hungarian is a left-headed language, thus, by its position, phrase-initial pitch accent marks the left “head” of the given accentual (minor) phrase. Accent appears on every word of the given structure: each main accent is equally “strong” [12], while intensity decreases during the sentence, and it “steps up” again at the beginning of a new intonational phrase (e.g. [13], [14]).

Therefore, the aim of the present study is to examine how boundaries of different strengths based on the variously coordinated names are realized in Hungarian. The main question of the examination is how the prosodic boundary marker patterns differ from each other regarding their position and syntactic structure (left or right-headed structures).

Our hypothesis: We hypothesized that left-headed structures would be more consequently (using more prosodic cues showing more notable changes) marked than right-headed ones.

2. Methodology

2.1 Material

A name-grouping task was conducted by adapting Wagner’s methodology [2] from English to Hungarian. The material consisted of differently coordinated ordinary, Hungarian first names. In the first group three, while in the second group four names were used in a row. All of them were disyllabic, and contained only short vowels and sonorant consonants (e.g. ‘Vili’, ‘Mira’, ‘Lili’, ‘Vali’ etc.). All the names were conjuncted with an ‘and’ (‘és’, &) or an ‘or’ (‘vagy’, /) conjunctor. The relation between the names were signed by using brackets, as well as appeared in a written form. (The results are presented without the symbols for simplicity.) The structures were embedded into a question-answer context. For the recordings, the SpeechRecorder program [15] was used. The test material started after a short trial part. Firstly, the question appeared on the screen that was always constant: ‘Ki menjen Gergővel moziba?’ - ‘Who should go to the cinema with Gergő?’ Then the participants saw the answer (once at a time). As a reference, the Name1 / Name2 / Name3 (/ Name4) structure was used without any grouping within the construction. In addition, seven differently coordinated structures were involved in the project (see 1. Table).

All structures were used in four variations with different names that were revealed in random order. The recordings were made in lab circumstances in the phonetic lab of the Hungarian Research Centre for Linguistics, Budapest using head microphones with omni-directional head microphone.

<table>
<thead>
<tr>
<th>Table 1: Structure forms used in the experiment.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three-name structures</td>
</tr>
<tr>
<td>A/B/C (ref.)</td>
</tr>
<tr>
<td>(A &amp; B) / C</td>
</tr>
<tr>
<td>A / (B &amp; C)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
2.2 Speakers
Altogether 15 monolingual Hungarian speakers participated in the study, who were recorded individually. Participants reported no known speech, hearing or language problems. They were instructed to read aloud the structures in a way to indicate which elements belong together.

Altogether, 540 structures were recorded and analyzed (15 subjects x 4 repetitions x 9 structures).

2.3 Analysis
Recordings were annotated firstly using the MAUS online annotator program [17], then corrected manually in Praat [18] on word level: every name as well as every conjunction were labeled. Those parts of the recordings that contained creaky voice, were marked manually, as well. The following acoustic parameters were analyzed in the study:

- duration of the names,
- mean fundamental frequency of the names (settings: time step: 0.01s, voiced/unvoiced cost: 0.3, pitch floor-ceiling: 75-300 Hz (men), 110-450 Hz (women) - individual modifications after manual control),
- appearance of silent pause after the names,
- mean intensity of the names (setting: time step: 0.01s + individual settings),
- appearance of creaky voice.

To eliminate the differences caused by the different articulation rate, relative word duration was calculated for each utterance: the duration of each name was divided by the mean name-duration of the utterance. So if the value is greater than 1, that means the actual word’s duration was greater than the mean name-duration at that utterance. Thus, if the value is lower than 1, the actual word duration was lower than the mean name-duration of the given utterance. The same calculation method was applied to the intensity and fundamental frequency (in semitones) values.

Regarding the small number of speakers, no statistical analysis was carried out.

3. Results

3.1 Duration
In the control group, a decreasing tendency was found in the duration of the names of the 3-element structures (ABC) (Fig. 1a), while firstly an increasing, then a decreasing tendency has been observed in the case of the 4-element (ABCD) ones (Fig. 1b). Compared to the reference group, a greater increase was realized before the boundaries in the other, differently coordinated structures, but no elongation has been found before the end of the structures. In addition, the elements occurred in the same group, conjuncted with “and” were realized with about the same relative duration.

In the case of 3-element structures longer duration was measured right before the boundaries compared to the reference values, but the ratio of elongation varied within the two structures. In the case of the left-headed ((ABC) units B was 5% longer, while in the right-headed structure (A(BC)) A was 15% longer compared to names in the same positions in the reference structure.

3.2. Fundamental frequency
In the reference group, irrespectively from the number of elements, a decreasing trend was observed towards the end of the structure (Figure 2a,b).
In the differently coordinated structures, a slightly increasing tendency occurred right before the boundaries regardless of their types.

In the case of the 3-element structures, lower mean f0 was found on the name before ‘and’ (at non-boundaries), and higher mean f0 realized on the name before ‘or’ (at boundaries). A more notable difference has been observed regarding the f0 characteristics of the left-headed structure ((AB)C) compared to the right-headed one (A(BC)): f0 increased with a larger extent (on average by 137% compared to the reference value) before the ‘or’ conjunctor in the right-headed structure (where there was no difference between the values of the reference and the given structure). Moreover, the f0-lowering before ‘and’ (at non-boundaries) was also larger in the left-headed structure (on average 51%), than in the right-headed one (on average 26%).

In the case of the 4-element structures, compared to the control group, the largest difference was observed in the (AB)(CD) structure. In the case of the other constructions, a decreasing trend occurred in the f0 mean values for each name, while in the (AB)(CD) construction an increase was described before the boundary marking preceding the second name (B). The more complex structures (structures containing one more embedding/boundary) do not differ notably from the simpler structures: c.f. (AB)CD vs. ((AB)C)D or AB(CD) vs. A(B(CD)). In the case of left-headed structures ((AB)CD, ((AB)C)D), the f0 mean of the first three names have been realized as closer values to each other’s ((AB)CD: 0.867±0.307; ((AB)C)D: 0.813±0.667 semitones) compared to the reference (0.905±1.176 semitones), were found to be more different from the f0 values of the four-member elements. In the case of right-headed structures (AB(CD), (A(B(CD))). Additionally, the average f0 of the four names were more distinct than in the left-headed structures (AB(CD): 0.772±1.094; (A(B(CD)): 0.862±1.198 semitones).

### 3.3. Silent pauses

In the control group irrespective of the number of names, silent pauses occurred in almost equal proportions (in the 3-element structures: after A: 46%, after B: 61%, in the 4-element struct.: after A: 54%, after B: 61%, after C: 57%). The roughly 50% ratio is mainly explained by individual differences: some speakers did not use the silent pause as a boundary marker at all in the reference structures, while others produced pauses in all relevant positions.

In the case of the 3-element structures both the left- and right-headed structure differed considerably from the reference structure (Fig. 3a): before the ‘or’ conjunctor, the ratio of the silent pauses was higher ((AB)C): 95%, A(BC): 79%), while before the ‘and’ conjunctor, that ratio was lower than in the reference structure ((AB)C): 0%, A(BC): 4%). A slightly larger difference was observed in the case of the left-headed constructions ((AB)C).

Regarding the 4-element structures (Fig. 3b), the given ratio before the ‘and’ (in non-boundary position, A&B) was lower than in the reference (between: 0-11%). In contrast, a higher silent pause ratio (compared to the reference structure) occurred in the other speech units: In the case of the simpler structures (AB(CD), (AB)CD) more silent pauses were found next to the boundaries (AB(CD): 46% < AB(CD): 86%, (AB)CD: 95% > (AB)CD: 71%). In addition, the two types of ‘or’ conjunctors were marked by silent pauses differently: the one that links the embeddings together appeared to be a stronger boundary since silent pauses appeared in a greater ratio. The tendency varied in the more complex structures: a slightly higher ratio was found in the structures with ‘and’ conjunctor, than for the boundaries with ‘or’ (((AB)C)D: 89% > (AB)C)D: 82%, A(B(CD)): 68% > A(B(CD)): 53%). The frequency of the silent pauses was moderately higher in the left-headed structures ((AB)CD, ((AB)C)D) than in the right-headed ones (89%, 82% > 68%, 53%). The usage of silent pauses was unambiguous in the case of the (AB)(CD) structures: the SP ratio was not higher than 7% before ‘and’ (A&B: 0%, C&D: 7%), and not lower than 90% before ‘or’ conjunctions (95%).
3.4. Intensity

A decreasing tendency has been observed regarding the mean intensity values of the names in the reference structures. In the case of 3-element structures the mean intensity of the name before ‘and’ conjunction was higher (17%) than in the reference group (Fig 4a). Thus a similar tendency to the reference was observed in left-headed structures ((AB)C), while an opposed tendency can be found at the right-headed ones (A(BC)): the mean intensity of the second name is higher than the first one.

In the case of the 4-element units, the right-headed simpler structure (AB(CD)) differed the least from the reference (Fig. 4b): for names that are linked by an ‘and’ conjunction, the largest change (26%) was measured in this case.

3.5. Creaky voice

Creaky voice seemed not to depend on the structure of the coordinated names: it appeared tendentiously more frequently at the end of the structures (in the last name) independently from the construction of the speech units. Its frequency varied between 60-100% in the last names, while it realized below 40% in most of the cases in other positions. The results showed remarkable individual differences.

3.6. Combination of the boundary markers

The combination of the appearance of lengthening, f0 rise and silent pause were analyzed as well (Fig. 5).

Irrespective of the structure, there the least boundary markers appeared in the last word of the structure (none, marked with pink), and most frequent boundary marker was lengthening in this position. Lengthening appeared more frequently at the end of the left-headed structures (e.g. (AB)C, (AB)CD), then in the right-headed ones. Reference structures had more boundary markers after the first word, and towards the end of the structure the ratio of the appearance of the boundary markers decreased.

In the case of 3-element structures all of the three boundary markers appeared mostly before the ‘or’ conjunction, besides, more boundary markers were found in the right-headed units (A(BC) > (AB)C).

The 4-element structures (except (AB)(CD)) contained two boundaries: In all cases all the three analyzed boundary markers occurred at both boundaries, but they were more frequent at the first one (AB(CD), A(B(CD)), (AB)CD, ((AB)CD).
Boundary markers were found slightly more frequently at right-headed structures, but mainly in the case of the symmetric structure ((AB)(CD)).

4. Conclusions and discussion

The aim of the present study was to investigate the realization of the prosodic boundary markers based on the read forms of differently grouped names firstly in Hungarian.

The results of the study confirmed a longer word duration before the appearance of the boundaries, presumably as an effect of the phrase final lengthening. In the case of structures containing two boundaries, the co-occurrence of several boundary markers was more common than that of single boundary markers. The probable hierarchy of boundary marker occurrence based on the results obtained is as follows: silent pause > phrase final lengthening > f0 increase > intensity change. The appearance of each of the prosodic boundary markers, based on the Proximity and Similarity model, indicate the distance between different elements that provide the basis for grouping based on the syntactic structure of the utterance.

The strongest boundary marking (the most significant difference from the control structure) was observed for the (AB)(CD) structure. Assumably, this is due to its simple and balanced construction, which is also used frequently in everyday communicative situations.

In contrast, for a large proportion of the grouped elements, no boundary marker appeared, and if so, only one element separately, which, based on the earlier mentioned model, corroborates the interrelation of the elements based on their proximity and similarity. The occurrence of multiple boundary markers was more typical in structures where the related elements were located on the left side (left-headed structures). It is likely to be explained by the characteristics of the Hungarian language and the chosen methodology (the structures were read separately, thus no boundary marking between the structures is observed).

Based on the examination of the similarities and differences between the languages, the results of the research contribute to the teaching of Hungarian as a foreign language and to the more accurate operation of the text-to-speech process.

5. Acknowledgements

The research was supported by the Alexander von Humboldt Foundation.

6. References